Journal of Mathematics, Computations, and Statistics (hal. 98 – 105)
Vol. 5. No. 2, Oktober 2022
http://www.ojs.unm.ac.id/jmathcos

Identifikasi Sebaran Karakteristik Kriminal di Indonesia Tahun 2021 Menggunakan *Model-Based Clustering*

Debora Chrisinta[1, a)], Leonard Peter Gelu [1, b)], dan Budiman Baso[1]

[1] *Program Studi Teknologi Informasi, Fakultas Sains dan Teknik, Universitas Timor*

[a)] deborachrisinta@unimor.ac.id
[b)]petergelu@unimor.ac.id

Abstrak. *Kriminalitas merupakan suatu aspek yang mempengaruhi kelancaran perekonomian dalam masyarakat. Pentingnya peranan pemerintah dalam meminimalisir terjadinya kriminalitas dapat dilakukan dengan mengetahui sebaran karakteristik kriminalitas yang ada di setiap provinsi. Metode Model-Based Clustering dapat membantu mengidentifikasi karakteristik tersebut. Proses Clustering dilakukan dengan memastikan terlebih dahulu sebaran peubah mendekati sebaran normal dilihat dari QQplot dan kebebasan antar peubah. Hasil Clustering menunjukkan bahwa terdapat dua karakteristik kriminalitas yang tersebar di provinsi Indonesia. Identifikasi karakteristik sebaran kriminalitas menggunakan nilai rata-rata pada masing-masing cluster optimal yang sudah didapatkan. Cluster pertama menunjukkan sebaran 16 provinsi dengan kategori kriminalitas cenderung lebih rendah, sedangkan cluster kedua menunjukkan sebaran 18 provinsi dengan kategori kriminalitas tinggi.*
Kata Kunci: *Kriminalitas, Karakteristik, Model-Based Clustering .*

Abstract. *Crime is an aspect that affects the smooth running of the economy in society. The importance of the government's role in minimizing the occurrence of crime can be done by knowing the distribution of criminal characteristics in each province. The Model-Based Clustering method can help identify these characteristics. The Clustering process is carried out by ensuring that the distribution of variables is close to the normal distribution as seen from the QQplot and the independence of variable. The results of the Clustering show that there are two characteristics of criminality that are spread across the province of Indonesia. Identify the characteristics of the distribution of crime using the average value in each of the optimal clusters that have been obtained. The first cluster shows the distribution of 16 provinces with lower crime categories, while the second cluster shows the distribution of 18 provinces with high crime categories.*
Keywords: *Crime, Characteristics, Model-Based Clustering .*

PENDAHULUAN

Salah satu aspek terciptanya keamanan dan stabilitas suatu negara dilihat berdasarkan tingkat keamanan yang ada pada negara tersebut. Indonesia termasuk salah satu negara urutan ke-25 yang memiliki tingkat kriminalitas cukup tinggi dibandingkan negara lainnya dalam skala global (Shaw, 2021). Peran pemerintah dalam menanggulangi masalah ini sangat diperlukan perhatian khusus, mengingat bahwa salah satu yang mempengaruhi berjalannya perekonomian suatu negara yaitu terjaminnya keamanan. Hal tersebut secara tidak langsung akan berdampak bagi suasana yang tercipta dalam lingkungan masyarakat, sehingga dapat menjalankan aktivitas perekonomian tanpa adanya ancaman akan keamaan (Wahyu, 2021). Adapun untuk meminimalisir kriminalitas yang ada di Indonesia, dapat diupayakan dengan memperoleh gambaran umum karakteristik masing-masing kriminalitas yang ada pada tiap provinsi. Hal ini dapat dilakukan dengan

melakukan identifikasi karakteristik dengan merapkan metode yang tersedia, sala satunya adalah *Model-Based Clustering*.

Model-Based Clustering adalah salah satu metode pengelompokan berbasis model yang biasa digunakan untuk mengidentifikasi karakteristik data yang berasal dari ada populasi berdasarkan distribusi probabilitas finite mixture (McNicholas, 2016). Beberapa peneliti sebelumnya telah mengevaluasi kinerja metode *Model-Based Clustering*, yang menunjukkan bahwa metode ini cenderung memberikan kinerja yang lebih baik dibandingkan metode yang *cluster* yang umum.

Greve *et.al* (2014) dalam hasil penelitiannya menunjukkan bahwa *Model-Based Clustering* memberikan performa kinerja yang lebih baik pada simulasi data dibandingkan dengan *Kmeans*. Meila dan Heckerman (2021) menunjukkan hasil yang sama dimana *Model-Based Clustering* memberikan keunggulan dalam kualitas hasil pengelompokan dibandingkan dengan metode *Kmeans*. Berdasarkan hasil penelitian tersebut, maka pada penulisan ini akan menerapkan metode *Model-Based Clustering* dalam mengidentifikasi sebaran karakteristik kriminalitas di Indonesia berdasarkan data Statistik Kriminalitas tahun 2021. Objek pengamatan dalam penelitian ini adalah provinsi yang ada di Indonesia, untuk selanjutnya di lakukan penerapan metode *Model-Based Clustering* untuk di identifikasi karakteristik sebaran kriminalitas.

METODE PENELITIAN

Variabel Penelitian

Adapun dalam proses identidikasi karakteristik kriminalitas yang ada di Indonesis maka dalam penelitian ini terdiri dari delapan variabel yang diamati. Variabel penelitian ditentukan berdasarkan data klasifikasi kejahatan di Indonesia tahun 2021. Berikut delapan variabel yang digunakan:

TABEL 1. Variabel Penelitian

Variabel	Keterangan
X_1	Kejahatan terhadap Nyawa
X_2	Kejahatan terhadap Fisik/ Badan
X_3	Kejahatan terhadap Kesusilaan
X_4	Kejahatan terhadap Kemerdekaan Orang
X_5	Kejahatan terhadap Hak Milik/ Barang dengan Penggunaan Kekerasan
X_6	Kejahatan terhadap Hak Milik/ Barang
X_7	Kejahatan Terkait Narkotika
X_8	Kejahatan terkait Penipuan, Penggelapan dan Korupsi

Jenis dan Sumber Data

Sumber data yang digunakan dalam penelitian ini berasal dari data sekunder yaitu data Statistik Kriminal 2021 (BPS, 2021). Data sekunder merupakan data yang dipublikasikan oleh lembaga pengumpul data resmi kepada masyarakat sebagai pengguna data (Purwanti, 2019).

Metode Analisis

1. *Model-Based Clustering*

Analisis *cluster* adalah salah satu teknik analisis dalam statistika yang digunakan untuk mengelompokkan objek berdasarkan karakteristik yang sama. Karakteristik objek yang digunakan dalam analisis *cluster* berasal dari variabel. Variabel dalam analisis *cluster* harus saling bebas dan tidak memiliki korelasi satu dengan yang lain (Nurhasanah *et al*, 2021). Salah satu

metode dalam analisis *cluster* adalah *Model-Based Clustering* yang diperkenalkan pertama kali oleh Banfield & Raftery (1993). Metode ini dalam pengelompokan objek pada mempunyai paremeter yang berbeda, dimana populasi dapat diperoleh dari subpopulasi yang memiliki distribusi tertentu. Sebuah peubah acak X yang berasal dari distribusi campuran memiliki fungsi kepadatan peluang sebagai berikut:populasi. Asumsi pada metode ini adalah subpopulasi

$$f(x|\vartheta) = \sum_{g=1}^{G} \pi_g f_g(x|\theta_g) \quad (1)$$

dimana $\pi_g > 0$ sehingga $\sum_{g=1}^{G} \pi_g = 1$ disebut sebagai proporsi campuran, $f_g(x|\theta_g)$ adalah komponen kepadatan ke-g, sedangkan ϑ dan π merupakan parameter.

2. Tahapan Analisis

Proses tahapan analisis data dilakukan dalam beberapa tahapan, berikut adalah tahapan dalam menerapkan metode *Model-Based Clustering* pada data Kriminalitas di Indonesia Tahun 2021:

(a). Eksplorasi data;
(b). Transformasi data;
(c). Menerapkan *Model-Based Clustering* ;
(d). Menentukan *cluster* optimal; dan
(e). Indentifikasi karakteristik berdasarkan nilai rataan tertinggi setiap *cluster* optimal yang telah terbentuk.

HASIL DAN PEMBAHASAN

Kejahatan merupakan kesalahan publik yang melakukan tindakan pelanggaran yang melanggar hukum negara dan sangat tidak disetujui oleh masyarakat. Kejahatan didefinisikan sebagai perbuatan atau kelalaian yang dilarang oleh undang-undang yang dapat dipidana dengan pidana penjara atau denda. Kategori tindakan kriminalitas diantaranya adalah pembunuhan, perampokan, perampokan, pemerkosaan, mengemudi dalam keadaan mabuk, penelantaran anak dan kegagalan untuk membayar pajak adalah contoh kejahatan. Istilah kejahatan berasal dari kata Latin "crimen" yang berarti pelanggaran dan juga pelaku kesalahan (Sowmyya, 2014). Beberapa klasifikasi kriminalitas yang terjadi di Indonesia disajikan pada Tabel 2.

TABEL 2. Klasifikasi Kejahatan di Indonesia

No	Klasifikasi Kejahatan	Jenis Kejahatan
1	Kejahatan terhadap nyawa	Pembunuhan
2	Kejahatan terhadap Fisik/ badan	Penganiayaan berat/ringan, kekerasan dalam rumah tangga
3	Kejahatan terhadap kesusilaan	Perkosaan dan pencabulan
4	Kejahatan terhadap Kemerdekaan Orang	Penculikan, kempekerjakan anak di bawah umur
5	Kejahatan terhadap hak milik/barang dengan penggunaan kekerasan	Pencurian dengan kekerasan, pencurian dengan kekerasan, menggunakan senjata api (Senpi), pencurian dengan kekerasan, menggunakan senjata tajam (Sajam)
6	Kejahatan terhadap hak milik/barang	Pencurian, pencurian dengan pemberatan, pencurian kendaraan bermotor, pengrusakan/penghanc-uran barang, pembakaran dengan sengaja dan penadahan
7	Kejahatan Terkait Narkotika	Narkotika dan Psikotropika

8	Kejahatan terkait penipuan, penggelapan dan korupsi	Penipuan/perbuatan curang, penggelapan dan korupsi

Sumber: Kitab Undang-Undang Hukum Pidana (KUHP) Republik Indonesia dan *The International Classification of Crime for Statistical Purposes* (ICCS) dalam Statistika Kriminal 2021

Identifikasi karakteristik kriminalitas di Indonesia diawali dengan melakukan analisis data awal untuk melihat gambaran umum dan selanjutnya melakukan uji korelasi antar variabel yang bertujuan untuk melihat kebebasan antar variabel sehingga tidak mempengaruhi kinerja metode *Clustering*. Setelah ditemukan tidak adanya keterikatan antar variabel maka selanjutnya menerapkan metode *Clustering*.

1. Gambaran Umum Statistik Kriminal di Indonesia Tahun 2021

Hasil dari perhitungan statistika deskriptif diperoleh bahwa variabel kejahatan terhadap hak milik/barang yang meliputi pencurian, pencurian dengan pemberatan, pencurian kendaraan bermotor, pengerusakan/ penghancuran barang, pembakaran dengan sengaja dan penadahan, merupakan tingkat kejahatan yang paling tinggi sering terjadi. Ha ini ditandai dengan nilai rata-rata tertinggi. Namun pada provinsi yang memiliki nilai maksimum kejahatan tersebut berada di provinsi Sumatera Utara yang ditandai dengan kategori kejahatan terhadap fisik dan badan (X2). Selanjutnya terdapat provinsi yang memiliki tingkat kejahatan yang tidak ada terjadi, yaitu provinsi Kalimantan Selatan. Berikut disajikan dalam Tabel 4 keseluruhan gambaran umum variabel pada statistik kriminalitas yang ada di Indonesia tahun 2022.

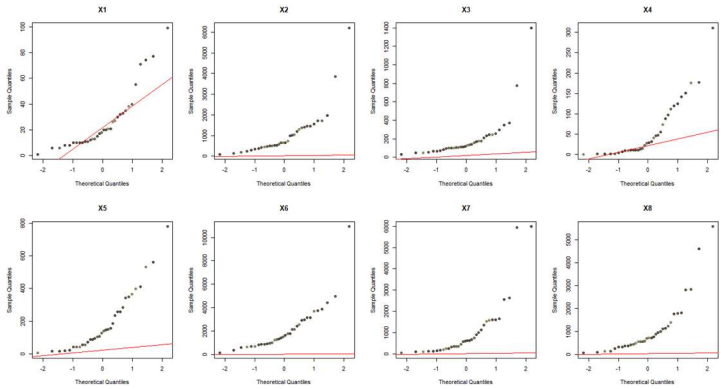
TABEL 3. Statistik Deskriptif Kriminal di Indonesia Tahun 2021

Variabel	Rataan	Varian	Max	Min
X1	26,4	548,1	99,0	1,0
X2	1078,6	1353960,3	6207,0	91,0
X3	202,1	63433,9	1398,0	32,0
X4	57,9	5007,0	311,0	0,0
X5	192,3	34040,4	780,0	4,0
X6	2154,8	3971480,3	10916,0	112,0
X7	1076,8	2000602,5	5981,0	41,0
X8	1091,1	1509017,0	5562,0	49,0

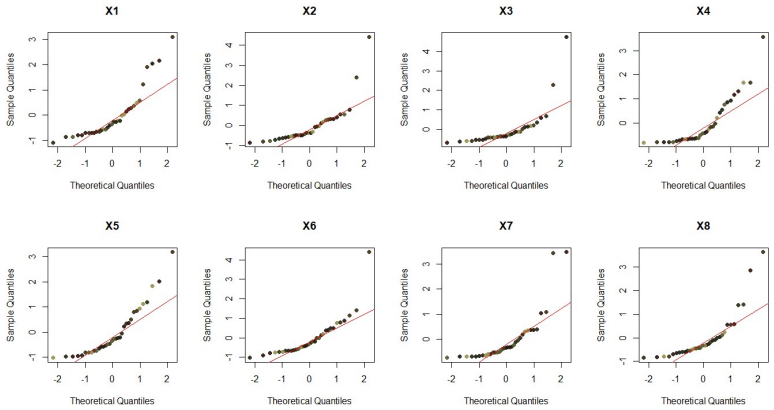
Berdasarkan Tabel 4 dapat dilihat pula untuk variabel X6 menunjukkan keberagaman data yang paling tinggi, hal ini menandakan bahwa kejahatan terhadap hak milik/barang yang terjadi di setiap provinsi memiliki jumlah kejahatan dengan selisih sebaran data yang cenderung tinggi dari satu provinsi ke provinsi yang lainnya. Adapun hal ini menandakan bahwa kejahatan yang terhadap hak milik/barang yang terjadi di setiap provinsi dapat dikatakan jarang terjadi di beberapa provinsi.

Proses identifikasi selanjutnya untuk mengetahui karakteristik sebaran kriminal di Indonesia Tahun 2021 diawali dengan melihat pola masing-masing kejahatan. Adapun metode *Model-Based Clustering* adalah metode berbasis sebaran data. Umumnya data yang ada dalam kejadian sehari-hari terdistribusi normal. Oleh karena itu, identifikasi sebaran jenis-jenis kejahatan dilakukan dengan melihat melalui grafik Q-Q Plots (Gambar 1). Q-Q Plots adalah grafik yang dibuat dengan memplot dua set kuantil terhadap satu sama lain. Jika kedua himpunan kuantil berasal dari distribusi yang sama, maka akan terlihat titik-titik yang membentuk garis linear (Andersen dan Dennison, 2018). Hasil Q-Q Plots menunjukkan bahwa titik-titik data tidak memencar disekitar garis lurus linear, sehingga data tidak menunjukkan terdistribusi normal. Oleh karena itu diperlukan transformasi data, sekaligus digunakan untuk meniadakan korelasi antar peubah. Hasil transformasi data ditunjukkan pada Gambar 2, yang menandai bahwa data telah terdistribusi normal. Hal tersebut dapat dilihat dari titik pencar data menyebar disekitar garis liner.

Selanjutnya, menangani korelasi data yang relatif tinggi dilakukan transformasi *Principal Component Analysis* (PCA) (Palkhouskaya *et al*, 2019), yang menunjukkan bahwa korelasi data antar variabel telah menunjukkan sebesar 0. Artinya bahwa tidak ada korelasi antar variabel.



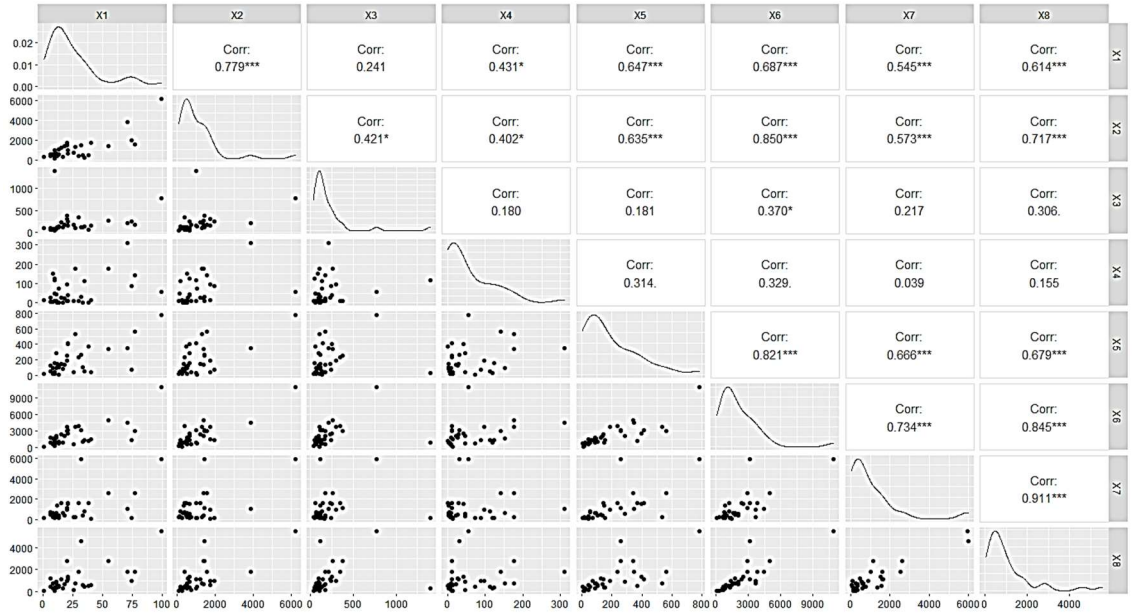
GAMBAR 1. Sebaran Data Statistik Kriminal di Indonesia Tahun 2021



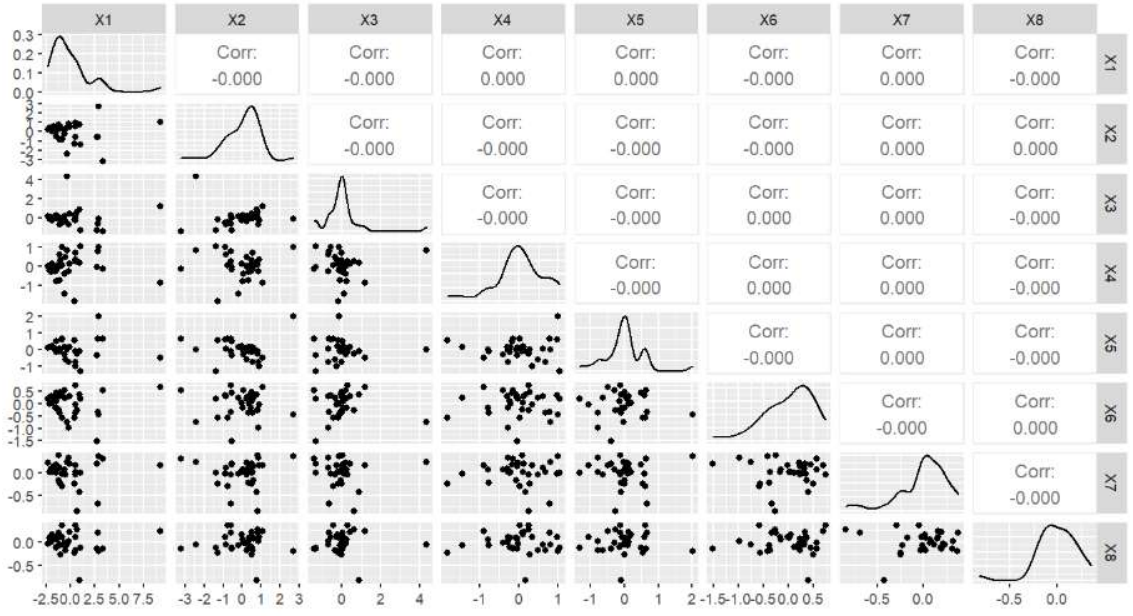
GAMBAR 2. Sebaran Data Statistik Kriminal di Indonesia Tahun 2021 Setelah Transformasi Data

Berikut hasil sebelum dan sesudah transformasi PCA untuk meniadakan korelasi antar variabel, dapat dilihat pada Gambar 3 dan 4. Sebelum dilakukan transformasi PCA ditemukan bahwa antar variabel memiliki korelasi yang sangat tinggi. Korelasi tertinggi ditemukan pada variabel X7 dan X8 yang merupakan kejahatan terkait narkotika dan kejahatan terkait penipuan, penggelapan dan korupsi.

Demikian untuk mayoritas variabel yang lainnya juga memiliki korelasi yang relatif tinggi. Variabel yang memiliki korelasi relatif kecil ditemupan pada variabel X4 dan X7 yang merupakan kejahatan terhadap kemerdekaan orang dan kejahatan terkait narkotika. Kondisi variabel yang tinggi akan menyebabkan pada saat menerapkan metode pengelompokkan memberikan informasi yang cenderung bias.



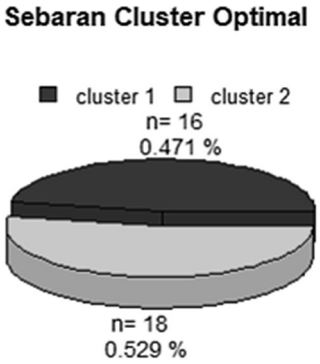
GAMBAR 3. Korelasi Variabel Kriminal di Indonesia Tahun 2021 Sebelum Transformasi Data



GAMBAR 4. Korelasi Variabel Kriminal di Indonesia Tahun 2021 Sesudah Transformasi Data

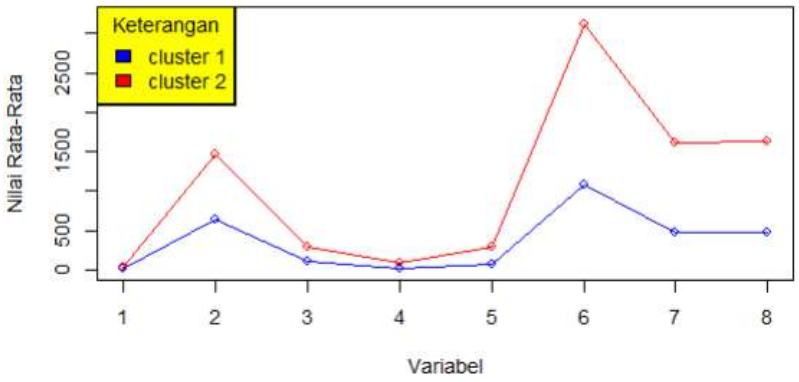
2. Hasil *Clustering*

Proses berikutnya setelah diperoleh data yang memenuhi untuk dilakukan penerapan metode *Model-Based Clustering* maka mencari *cluster* optimal untuk menggambarkan karakteristik sebaran kriminal di Indonesia. *Cluster* optimal diperoleh sebanyak 2 *cluster*, sebaran banyaknya provinsi yang masuk dalam kedua *cluster* dapat dilihat pada Gambar 5.



GAMBAR 5. Sebaran Jumlah Provinsi berdasarkan *Cluster Optimal*

Sebanyak 16 provinsi berada di *cluster* 1 dan 18 provinsi berada di *cluster* 2. Adapun karakteristik dari masing-masing *cluster* dilihat berdasarkan nilai rata-rata masing-masing peubah disajikan pada Gambar 6. Terdapat 18 provinsi yang memiliki tingkat kriminal yang tinggi berada pada *cluster* 2. Sedangkan 16 provinsi lainnya yang berada pada *cluster* 1 memiliki tingkat kriminal yang cenderung lebih rendah. Adapun pada *cluster* 2 tingkat kejahatan terhadap hak milik/barang adalah yang paling sering terjadi, kriminalitas dalam kategori ini antara lain mencakup Pencurian, pencurian dengan pemberatan, pencurian kendaraan bermotor, pengrusakan/penghancuran barang, pembakaran dengan sengaja dan penadahan.



GAMBAR 6. Sebaran Jumlah Provinsi berdasarkan *Cluster Optimal*

Berikut adalah provinsi yang termasuk pada masing-masing *cluster* optimal:

TABEL 4. Sebaran Provinsi di Indonesia berdasarkan *Cluster* Optimal

No	*Cluster* 1	*Cluster* 2
1	Aceh	Banten
2	Bali	Jawa Barat
3	Bengkulu	Jawa Tengah
4	DI Yogyakarta	Jawa Timur
5	Gorontalo	Kalimantan Barat
6	Jambi	Kalimantan Selatan
7	Kalimantan Tengah	Lampung
8	Kalimantan Timur	Maluku
9	Kalimantan Utara	Metro Jaya

10	Kep. Bangka Belitung	Nusa Tenggara Barat
11	Kepulauan Riau	Papua
12	Maluku Utara	Papua Barat
13	Nusa Tenggara Timur	Riau
14	Sulawesi Barat	Sulawesi Selatan
15	Sulawesi Tengah	Sulawesi Utara
16	Sulawesi Tenggara	Sumatera Barat
17		Sumatera Selatan
18		Sumatera Utara

KESIMPULAN

Berdasarkan metode *Model-Based Clustering* yang digunakan untuk mengidentifikasi sebaran karakteristik kriminal di Indonesia tahun 2021 diperoleh *cluster* optimal sebanyak 2 *cluster*. *Cluster* 1 berisi 16 provinsi dengan kategori kriminalitas cenderung lebih rendah, sedangkan custer 2 berisi 18 provinsi dengan kriminalitas cenderung tinggi.

DAFTAR PUSTAKA

Andersen AJ & Dennison JR. (2019). An Introduction Quantile-Quantile Plots for the Experimental Physicist. Am. J. Phys, 87 (5).

BPS. (2021). Statistik Kriminal. Jakarta.

Greve B, Pigeot I, Huybrechts I, Pala V & Börnhorst C. (2014). A Comparison Of Heuristic And Model-Based Clustering Methods For Dietary Pattern Analysis. Public Health Nutrition, 19(2), 255–264.

McNicholas, PD. (2016). Model-Based Clustering. Journal of Classification, 33:331-373.

Meila M & Heckerman D. (2001). An Experimental Comparison of Model-Based Clustering Methods. Machine Learning. 42, 9–29.

Nurhasanah N, Salwa N, Ornila L, Hasan A & Mardhani M. (2021). Classifying Regencies And Cities On Human Development Index Dimensions: Application Of K-Means Cluster Analysis. Jurnal Sains Sosio Humaniora, Vol. 5 No 2, 913-918.

Palkhouskaya, et al. (2019). Method For Clustering Distributions Grid. INTEgrated opeRation PLAnning tool towards the Pan-European Network. Project co-funded by the European Commission within the H2020 Programme.

Purwanti EY & Widyaningsih E. (2019). Analisis Faktor Ekonomi Yang Mempengaruhi Kriminalitas Di Jawa Timur. Jurnal Enokomi, Vol. 9, No. 2.

Shaw M. (2021). Global Organized Crime Index. Enact : Institute For Security Studies And Interpol.

Wahyu AM, Anugrah PG, Danyalin AM, & Noorrizki RD. (2021). Ketimpangan Ekonomi Berdampak pada Tingkat Kriminalitas? Telaah dalam Perspektif Psikologi Problematika Sosial. Jurnal Ilmiah Ilmu Sosial, Vol. 7 No 2, 170-178.