TINGKAT KEPARAHAN DAN RISIKO PENYEBARAN COVID-19 DI INDONESIA DENGAN MENGGUNAKAN K-MEANS CLUSTERING

(Severity Level and Spreading Risk of Covid-19 in Indonesia by K-Means Clustering)

Diah Novita Sari¹, Irma Yunita²

Badan Pusat Statistik Kabupaten Bangka Barat¹ Badan Pusat Statistik Kabupaten Mesuji²

> Muntok, Bangka Barat E-mail: <u>diah.novita@bps.go.id</u>

ABSTRAK

Terhitung hingga September 2020, kasus positif Covid-19 di Indonesia terus menunjukkan kenaikan. Pada 15 September 2020, jumlah kasus terkonfirmasi positif di Indonesia mencapai 225.030 kasus dengan total kasus meninggal sebanyak 8.965 kasus. Sementara itu, terdapat perbedaan yang cukup signifikan pada kasus terjadinya Covid-19 di tiap provinsi, yang artinya tingkat keparahan tiap wilayah pun berbeda. Sehingga, beberapa provinsi yang memiliki tingkat keparahan tinggi seharusnya lebih berusaha untuk menaruh perhatian dan memberi perlakuan khusus terhadap wilayahnya. Oleh karena itu, penelitian ini bertujuan untuk mengetahui gambaran umum mengenai tingkat keparahan serta risiko penyebaran *covid*-19 pada 34 provinsi di Indonesia melalui *cluster* yang terbentuk. Analisis yang digunakan adalah *clustering analysis* dengan metode *K-Means*. Hasil pengujian menunjukkan bahwa DKI Jakarta dan Jawa Timur mempunyai tingkat keparahan terhadap *covid*-19 tertinggi di Indonesia.

Kata kunci: covid-19, k-means clustering, tingkat keparahan, risiko penyebaran

ABSTRACT

As of September 2020, positive cases of Covid-19 in Indonesia continue to show an increase. On September 15, 2020, the number of confirmed positive cases in Indonesia has reached 225,030 cases with the total number of cases died was 8,965. Meanwhile, there are significant differences in the cases of Covid-19 in each province, which means that the severity of each region is different. Thus, several provinces with a high level of severity should put more effort into paying attention and giving special treatment to their regions. Therefore, this study aims to find a general picture of the severity of Covid-19 in 34 provinces in Indonesia through the formed clusters. The analysis used is clustering analysis with the K-Means method. The test results show that DKI Jakarta and East Java have the highest severity levels of COVID-19 in Indonesia.

Keywords: covid-19, k-means clustering, severity level, spreading risk

PENDAHULUAN

Saat ini, dunia tengah dilanda pandemi bernama virus *corona* atau *covid-19*, tak terkecuali negara Indonesia. *Covid-*19 (*coronavirus disease* 2019) merupakan jenis penyakit baru yang disebabkan oleh virus dari golongan *coronavirus*, yaitu SARS-CoV-2 yang juga sering disebut virus *corona* (Lipsitch, Swerdlow, dan Finelli; 2020). Virus ini mulai muncul pertama kali di Wuhan, Tiongkok pada bulan Desember 2019 dan mulai masuk ke Indonesia pada awal bulan Maret 2020.

Kemunculan virus *Covid*-19 di Indonesia tentu tidak dapat diabaikan. Berbagai upaya penanggulangan *Covid-19* telah dilakukan oleh pemerintah Indonesia guna menghadang laju

penyebaran, mulai dari penerapan *physical distancing* hingga Pembatasan Sosial Berskala Besar (PSBB) di berbagai daerah yang terpetakan sebagai episentrum penyebaran.

Terhitung hingga September 2020, kasus positif *Covid*-19 di Indonesia terus menunjukkan kenaikan. Terhitung sampai 15 September 2020, jumlah kasus terkonfirmasi positif di Indonesia mencapai 225.030 kasus. Sedangkan total kasus meninggal sebanyak 8.965 kasus. Tingginya jumlah kasus terkonfirmasi positif tersebut merupakan kumulatif dari jumlah kasus dari setiap provinsi di Indonesia, di mana terdapat perbedaan yang cukup signifikan di tiap provinsi, yang artinya tingkat keparahan tiap wilayah pun berbeda. Sehingga, beberapa provinsi yang memiliki tingkat keparahan tinggi seharusnya lebih berusaha untuk menaruh perhatian dan memberi perlakuan khusus terhadap wilayahnya. Sementara itu, perlu dilihat bagaimana tingkat risiko penyebaran *covid*-19 pada tiap provinsi di Indonesia. Hal ini merupakan salah satu upaya untuk menekan penyebaran *covid*-19 di Indonesia. Oleh karena itu, penelitian ini ingin melihat gambaran umum mengenai tingkat keparahan serta tingkat risiko penyebaran *Covid*-19 pada provinsi-provinsi di Indonesia melalui *cluster* yang terbentuk.

METODE

Penelitian ini bertujuan untuk mengetahui gambaran umum mengenai tingkat keparahan serta tingkat penyebaran *covid*-19 pada tiap provinsi di Indonesia melalui *cluster* yang terbentuk. Adapun *cluster* diperoleh dengan cara mengelompokkan provinsi-provinsi di Indonesia berdasarkan berbagai tingkat keparahan serta tingkat risiko penyebaran *covid*-19.

Untuk dapat memenuhi tujuan penelitian, maka digunakan metode analisis *clustering* dengan metode *K-Means*. Metode ini karena dianggap cocok untuk mengelompokkan provinsi di Indonesia berdasarkan kesamaan (*similarity*) dan perbedaan (*dissimilarity*) karakteristik, yaitu tingkat keparahan serta tingkat risiko penyebaran *Covid*-19 pada tahun 2020 di Indonesia.

Algoritma *K-means* merupakan metode *clustering* yang paling banyak digunakan. J. Macqueen awalnya mengusulkan algoritma *K-Means* pada tahun 1967. (Wagstaff, dkk.; 2001). *K-Means Cluster Analysis* merupakan salah satu metode analisis klaster non hirarki yang dapat digunakan untuk mempartisi objek ke dalam kelompok-kelompok berdasarkan kedekatan karakteristik, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu klaster yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokan ke dalam klaster yang lain. Tujuan pengelompokan adalah untuk meminimalkan *objective function* yang diset dalam proses pengelompokan, yang pada dasarnya berusaha untuk meminimalkan variasi dalam satu klaster dan memaksimalkan variasi antar klaster.

Jika diberikan sekumpulan objek X=X1,X2,...,Xn maka algoritma *K-Means Cluster Analysis* akan mempartisi X dalam k buah klaster, di mana setiap klaster memiliki *centroid* dari objek-objek dalam klaster tersebut. Pada tahap awal algoritma *K-Means Cluster Analysis* dipilih secara acak k buah objek sebagai *centroid*, kemudian jarak antara objek dengan *centroid* dihitung dengan menggunakan jarak *Euclidean*, objek ditempatkan dalam klaster yang terdekat dihitung dari titik tengah klaster. *Centroid* baru ditetapkan jika semua objek sudah ditempatkan dalam klaster terdekat. Proses penentuan *centroid* dan penempatan objek dalam klaster diulangi sampai nilai *centroid* konvergen (*centroid* dari semua klaster tidak berubah lagi) (Johnson dan Wichern, 2007).

Berdasarkan penelitian oleh COVID dan Team (2020), tingkat keparahan suatu penyakit diindikasikan oleh tiga hal yakni dirawat di rumah sakit, masuk ruang *Intensive Care Unit* (ICU), serta kasus kematian. Sementara pada penelitian yang dilakukan oleh Coşkun, Yıldırım, dan Gündüz (2020) menunjukkan bahwa penyebaran *Covid*-19 di Turki dipengaruhi oleh tingkat kepadatan penduduk di tiap daerah. Selain itu, penelitian oleh Copiello dan Grillenzoni (2020) juga menunjukkan bahwa penyebaran Covid-19 di China salah satunya disebabkan oleh tingginya tingkat kepadatan penduduk.

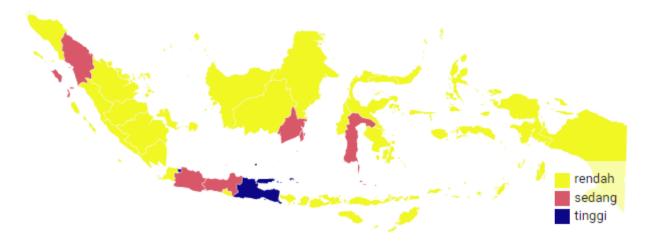
Penelitian ini mencakup 34 provinsi di Indonesia dengan variabel yang digunakan beserta sumber datanya sebagai berikut:

- Case Fatality Rate (CFR), dihitung dari jumlah kasus kematian dibagi dengan total kasus dikaitkan 100. Sumber data dari variabel ini diperoleh dari Kawal Covid-19 per 15 september 2020. CFR dihitung dengan membagi jumlah total kematian akibat penyakit dengan jumlah kasus yang dikonfirmasi. CFR digunakan sebagai ukuran keparahan suatu penyakit/wabah (Onder, Rezza, dan Brussaferro; 2020). CFR Covid-19 berbedabeda pada tiap wilayah dan dapat berubah seiring periode terjadinya wabah (Roser, dkk. 2020).
- 2. Total kasus, yaitu jumlah kasus terkonfirmasi positif *covid-19*. Sumber data dari variabel ini diperoleh dari Kawal *Covid-*19 per 15 september 2020.
- 3. Kepadatan penduduk, dengan sumber data proyeksi penduduk 2019 oleh BPS. Pembentukan klaster berdasarkan tingkat keparahan terjadinya *Covid*-19 dilakukan dengan menggunakan dua variabel, yakni variabel total kasus dan variabel CFR. Sementara pembentukan klaster tingkat risiko penyebaran *Covid*-19 dilakukan dengan menggunakan variabel total kasus dan kepadatan penduduk.

HASIL DAN PEMBAHASAN

Dalam penelitian ini, proses iterasi dilakukan sebanyak tiga kali. Proses ini dilakukan untuk mendapatkan *cluster* yang tepat. Dapat diketahui bahwa jarak minimum antar pusat *cluster* yang terjadi dari hasil iterasi adalah 17.839.

Berdasarkan pengelompokan tingkat keparahan *Covid-*19 dengan *K-Means Clustering*, terdapat tiga klaster yang terbentuk. Klaster ini terdiri dari tiga kelompok daerah dengan tingkat keparahan wabah *Covid-*19, yakni rendah, sedang, dan tinggi. Dengan demikian, dapat terlihat pada Gambar 1 pengelompokan daerah-daerah dengan tingkat keparahan tinggi, sedang, maupun rendah.



Gambar 1. Peta Sebaran Tingkat Keparahan Covid-19 Menurut Provinsi di Indonesia Tahun 2020.

Klaster 1 berisikan provinsi yang mempunyai rata-rata tingkat keparahan terhadap *covid*-19 berada di bawah rata-rata total, yaitu tampak berwarna kuning pada Gambar 1. Artinya, provinsi tersebut memiliki tingkat keparahan yang rendah terhadap *covid*-19, dilihat dari CFR dan total kasus. Berdasarkan hasil pengolahan, terdapat 27 provinsi yang tergolong ke dalam klaster ini, yaitu Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, dan Papua.

Klaster 2 yang tampak berwarna merah pada Gambar 1 berisikan provinsi yang mempunyai rata-rata tingkat keparahan terhadap *covid*-19 mendekati nilai rata-rata total. Artinya, provinsi

tersebut memiliki tingkat keparahan yang sedang terhadap *covid*-19, dilihat dari CFR dan total kasus. Terdapat lima provinsi yang tergolong ke dalam klaster ini, di antaranya Sumatera Utara, Jawa Barat, Jawa Tengah, Kalimantan Selatan, dan Sulawesi Selatan. Adapun DKI Jakarta dan Jawa Timur merupakan provinsi yang tergolong ke dalam klaster 3, yang tampak berwarna biru pada Gambar 1.

Klaster 3 merupakan klaster yang mempunyai rata-rata tingkat keparahan terhadap *covid*-19 berada di atas rata-rata total. Artinya, provinsi tersebut memiliki tingkat keparahan yang tinggi terhadap *covid*-19, dilihat dari CFR dan total kasus.

Berdasarkan Gambar 1 diatas, hampir seluruh provinsi di Pulau Jawa memiliki tingkat keparahan terhadap *Covid*-19 dengan kategori sedang dan tinggi. Sementara itu, Provinsi Bengkulu menjadi satu satunya provinsi dengan status keparahan Covid-19 tertinggi Pulau Sumatera. Di Pulau Kalimantan, Kalimantan selatan menjadi satu-satunya provinsi dengan tingkat keparahan *Covid*-19 tertinggi. Sementara Sulawesi Selatan, juga menjadi satu-satunya provinsi dengan tingkat keparahan Covid-19 tertinggi di Pulau Sulawesi. Sementara di wilayah Maluku maupun Papua, tingkat keparahan terhadap *Covid*-19 masih tergolong rendah.

Jumlah anggota tiap klaster pada pengelompokan berdasarkan tingkat keparahan *Covid*-19 dapat dilihat pada Tabel 1 berikut.

Tabel 1. Jumlah Anggota di Masing-Masing Klaster Berdasarkan Tingkat Keparahan *Covid*-19.

Klaster yang Terbentuk			
Klaster	1	27	
	2	5	
	3	2	

Adapun berdasarkan analisis Anova pada Tabel 2 berikut dapat diketahui bahwa variabel total kasus merupakan variabel yang paling menunjukkan adanya perbedaan di antara provinsi-provinsi pada ketiga klaster yang terbentuk. Hal ini ditunjukkan dengan nilai F = 201,553 dan nilai p-value = 0,000. Di mana semakin besar nilai F, maka semakin besar pula perbedaan variabel pada klaster yang terbentuk.

Tabel 2. Hasil Uji Anova Klaster Tingkat Keparahan Covid-19.

ANOVA						
	Cluster		Error		F	Sig.
			Mean			
	Mean Square	df	Square	df		
total_kas	2011022721.2	2	9978647.9	31	201.533	.000
us	35		35			
CFR	5.428	2	2.664	31	2.038	.147

Selanjutnya, pengelompokan berdasarkan tingkat risiko penyebaran *Covid*-19 diwakili oleh variabel kepadatan penduduk dan variabel total kasus *Covid*-19 yang terjadi di masing-masing provinsi. Variabel ini diklasifikasikan menjadi tiga klaster dengan tiga kali proses iterasi. Berdasarkan hasil pengolahan, Tabel 3 menunjukkan bahwa terdapat 32 anggota yang tergolong ke dalam klaster ini. Klaster pertama yang terbentuk terdiri atas 32 anggota. sementara klaster 2 dan 3 terdiri dari satu anggota.

Tabel 3. Jumlah Anggota Masing-Masing Klaster Berdasarkan Tingkat Risiko Penyebaran *Covid*-19.

Klaster yang Terbentuk			
Klaster	1	32	

2	1	
3	1	

Klaster 1 berisikan provinsi yang mempunyai rata-rata *risiko* penyebaran Covid-19 di bawah rata-rata total. Klaster 2 berisikan provinsi yang mempunyai rata-rata risiko terkena covid-19 mendekati nilai rata-rata total. Artinya, provinsi tersebut memiliki risiko yang sedang untuk terkena *covid*-19. Terdapat satu provinsi yang tergolong ke dalam klaster ini, yakni Jawa Timur. Adapun DKI Jakarta juga merupakan satu-satunya provinsi yang tergolong ke dalam klaster 3. Klaster 3 merupakan klaster yang mempunyai rata-rata risiko terkena covid-19 berada di atas rata-rata total. Artinya, provinsi tersebut memiliki risiko yang tinggi untuk terkena covid-19. Untuk lebih jelasnya, sebaran tingkat risiko penyebaran *covid*-19 dapat dilihat pada Gambar 2.



Gambar 2. Peta Sebaran Tingkat Risiko Penyebaran Covid-19 Menurut Provinsi di Indonesia Tahun 2020.

Berdasarkan Gambar 2 diatas, terlihat bahwa tingkat penyebaran *Covid*-19 tertinggi berada di Provinsi DKI Jakarta. Hal ini didukung bahwa DKI Jakarta merupakan pusat pemerintahan maupun pusat perekonomian di Indonesia. Sementara Jawa Timur, merupakan provinsi dengan penyebaran *Covid*-19 tingkat penyebaran *Covid*-19 tertinggi kedua setelah DKI Jakarta. Hal ini didukung bahwa Jawa Timur merupakan provinsi dengan tingkat kepadatan penduduk yang tinggi serta kegiatan perekonomian yang aktif.

Berdasarkan analisis Anova pada Tabel 4, dapat diketahui bahwa variabel kepadatan penduduk memiliki nilai F = 7.298,187 dan nilai p-value = 0,000. Sementara variabel total kasus memiliki nilai F = 7298.187 dan p-value = 0.000.

Tabel 4. Hasil Uji Anova Klaster Tingkat Risiko Penyebaran *Covid*-19.

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
kepadatan_p	118518325.5	2	164441.415	31	7298.187	.000
dd	1850450124	2	20338170.32	31	90.984	.000
total_kasus						

KESIMPULAN

Berdasarkan pengelompokan menurut tingkat keparahan kasus Covid-19 yang terjadi, provinsi yang berada di Klaster 1 dapat dikatakan sebagai provinsi dengan tingkat keparahan terhadap covid-19 rendah, sementara Klaster 2 dikatakan sebagai provinsi dengan tingkat keparahan terhadap *covid*-19 sedang, dan Klaster 3 dikatakan sebagai provinsi dengan tingkat keparahan tinggi terhadap *covid*-19. Terdapat 27 provinsi yang memiliki tingkat keparahan yang rendah terhadap *covid*-19, lima provinsi yang memiliki tingkat keparahan yang sedang terhadap covid-19, dan dua provinsi yang memiliki tingkat keparahan yang tinggi terhadap covid-19. Sementara itu, variabel total kasus merupakan variabel yang paling menunjukkan adanya perbedaan di antara provinsi-provinsi pada ketiga cluster yang terbentuk. Sedangkan berdasarkan pengelompokan menurut tingkat risiko penyebaran Covid-19, provinsi yang berada di Klaster 1 dapat dikatakan sebagai provinsi dengan tingkat risiko penyebaran covid-19 yang rendah, sementara Klaster 2 dikatakan sebagai provinsi dengan tingkat tingkat risiko penyebaran covid-19 sedang, dan Klaster 3 dikatakan sebagai provinsi dengan tingkat risiko penyebaran Covid-19 yang tinggi. Terdapat 32 provinsi yang memiliki tingkat keparahan yang rendah terhadap *covid*-19, satu provinsi yang memiliki tingkat keparahan yang sedang terhadap covid-19, dan satu provinsi yang memiliki tingkat keparahan yang tinggi terhadap covid-19.

Berdasarkan tingkat keparahan maupun tingkat risiko penyebaran Covid-19, DKI Jakarta menjadi satu-satunya provinsi yang memiliki tingkat keparahan maupun tingkat risiko penyebaran Covid-19 tertinggi di Indonesia. Selain itu, dengan tingkat keparahan maupun tingkat risiko penyebaran dengan kategori sedang, ditempati oleh Provinsi Jawa Timur. Hal ini menunjukkan bahwa perlu adanya intervensi pemerintah daerah, khususnya pada Provinsi DKI Jakarta maupun Provinsi Jawa Timur terkait *Covid*-19. Beberapa daerah di luar Pulau Jawa yang perlu menjadi perhatian oleh pemerintah diantaranya Provinsi Bengkulu, Provinsi Kalimantan Selatan, serta Provinsi Bali.

DAFTAR PUSTAKA

- Carozzi, F. (2020). Urban density and COVID-19.
- Copiello, S., & Grillenzoni, C. (2020). *The spread of 2019-nCoV in China was primarily driven by population density. Comment on "Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China" by Zhu et al. Science of The Total Environment,* 744, 141028. doi:10.1016/j.scitotenv.2020.141028
- Coşkun, H., Yıldırım, N., & Gündüz, S. (2020). *The spread of COVID-19 virus through population density and wind in Turkey cities*. Science of The Total Environment, 141663.
- COVID, T. C., & Team, R. (2020). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19)-United States, February 12-March 16, 2020. MMWR Morb Mortal Wkly Rep, 69(12), 343-346.
- Johnson, R. Dean Wichern., 2007. Applied Multivariat Statistical Analysis.
- Jordan, R. E., Adab, P., & Cheng, K. K. (2020). Covid-19: risk factors for severe disease and death.
- Lipsitch, M., Swerdlow, D. L., & Finelli, L. (2020). *Defining the epidemiology of Covid-19—studies needed*. New England journal of medicine, 382(13), 1194-1196.
- Lucas, C., Wong, P., Klein, J., Castro, T. B., Silva, J., Sundaram, M., ... & Takahashi, T. (2020). Longitudinal analyses reveal immunological misfiring in severe COVID-19. Nature, 584(7821), 463-469.
- Roser, M., Ritchie, H., Ortiz-Ospina, E., & Hasell, J. (2020). *Coronavirus disease (COVID-19)–Statistics and research*. Our World in data.
- Rajgor, D. D., Lee, M. H., Archuleta, S., Bagdasarian, N., & Quek, S. C. (2020). *The many estimates of the COVID-19 case fatality rate*. The Lancet Infectious Diseases, 20(7), 776-777.
- Velavan, T. P., & Meyer, C. G. (2020). *The COVID-19 epidemic*. Tropical medicine & international health, 25(3), 278.
 - Wagstaff, K., Cardie, C., Rogers, S., & Schrödl, S. (2001, June). *Constrained k-means clustering with background knowledge*. In Icml (Vol. 1, pp. 577-584).
- World Health Organization. (2020). Coronavirus disease 2019 (COVID-19): situation report, 72.

Seminar Nasional Official Statistics 2019: Pengembangan Official Statistics dalam mendukung Implementasi SDG's