http://jurnal.unsur.ac.id/JMTSI

e-issn: <u>2581-0561</u> p-issn: <u>2581-0529</u>

DOI: 10.35194/jmtsi.v5i1.1220

Analisis Karakteristik Wilayah Transmisi Covid-19 dengan Menggunakan Metode *K-Means Clustering*

Dina Tri Utari^{1*}

¹Program Studi Statistika Universitas Islam Indonesia Jl. Kaliurang Km 14.5 Sleman Yogyakarta ^{1*}dina.t.utari@uii.ac.id

Characteristics Analysis of Covid-19 Transmission Area Using the K-Means Clustering Method

Dikirimkan: 02, 2021. Diterima: 03, 2021. Dipublikasikan: 03, 2021.

Abstract— Since the beginning of 2020, Indonesia has become one of the countries affected by the Covid-19 pandemic. Various efforts have been made by the government to prevent wider disease transmission. Large-scale social restrictions are one of the efforts that have been made by the Government. Bali Province is one of the areas where there are quite a lot of community activities, considering that Bali is a tourist destination that is in great demand by local and foreign tourists. This study aims to see the incidence of positive cases of Covid-19 based on the type of Covid-19 transmission that has occurred in all areas of Bali, so that the mitigation design can be adjusted based on the characteristics of the source of infection in various existing areas. The results show that based on the transmission source, it can be grouped into four clusters that have their respective characteristics. The proposed mitigation strategies include restrictions on local transmission and domestic travel for areas in clusters 1, 2, and 3. Meanwhile, restrictions on local transmission and overseas travel are in the 4th cluster.

Keywords—covid-19, K-Means clustering, profiling, risk mitigation.

Abstrak— Sejak awal tahun 2020, Indonesia menjadi salah satu negara terkena pandemi Covid-19. Berbagai upaya telah dilakukan oleh pemerintah untuk mencegah transmisi penyakit yang lebih luas. Pembatasan sosial berskala besar menjadi salah satu upaya yang telah dilakukan oleh Pemerintah. Provinsi Bali merupakan salah satu wilayah yang cukup banyak terjadi aktivitas masyarakat, mengingat Bali merupakan kawasan destinasi wisata yang banyak diminati wisatawan lokal maupun mancanegara. Penelitian ini bertujuan untuk melihat kejadian kasus positif Covid-19 berdasarkan jenis transmisi Covid-19 yang terjadi di seluruh wilayah Bali, sehingga rancangan mitigasi dapat disesuaikan berdasarkan karakteristik sumber infeksi di berbagai wilayah yang ada. Hasil menunjukkan, bahwa berdasarkan sumber transmisi dapat dikelompokkan menjadi empat cluster yang memiliki karakteristik masing-masing. Usulan strategi mitigasi yang diberikan antara lain pembatasan transmisi lokal dan perjalanan dalam negeri untuk wilayah yang berada pada cluster 1, 2, dan 3. Sedangkan pembatasan transmisi lokal serta perjalanan luar negeri pada cluster ke-4.

Kata kunci— covid-19, *K-Means clustering*, profiling, mitigasi risiko

I. PENDAHULUAN

Sejak awal tahun 2020, Indonesia menjadi salah satu negara terkena pandemi Covid-19. Indonesia melalui Satuan Tugas Covid-19 berusaha untuk melakukan pendataan maupun mitigasi risiko penularan untuk mencegah terjadinya kasus yang lebih banyak.

Peningkatan kasus Covid-19 yang melebihi kapasitas sumber daya medis membuat penanganan kasus menjadi lebih lama. Hingga tanggal 9 September 2020, tercatat pasien yang masih dalam perawatan berjumlah 1.196 orang. Sementara, mereka yang masih mendapat perawatan di rumah sakit sebanyak 668 orang. Jumlah tempat tidur yang tersedia di 55 rumah sakit rujukan Covid-19 di Bali sebanyak 778 kasur. Dari jumlah kasus dan ketersediaan tempat tidur khusus pasien Covid-19, rumah sakit di Bali

terancam penuh jika terjadi kasus Covid-19 terus bertambah [1].

Keterbatasan kapasitas sumber daya medis tidak hanya pada peralatan medis dan isolasi atau ruang rawat inap, tetapi juga dari segi tenaga medis dan dokter. Tim Mitigasi Ikatan Dokter Indonesia (IDI) melaporkan pembaruan data kematian tenaga medis akibat Covid-19. Sejak Maret sampai akhir Desember 2020, sebanyak 504 tenaga medis wafat terinfeksi Covid-19. Terdapat 11 provinsi dengan kematian tenaga medis dan kesehatan tertinggi mencakup Jawa Timur, DKI Jakarta, Jawa Tengah, Sumatra Utara, Jawa Barat, Sulawesi Selatan, Banten, Bali, Nanggroe Aceh Darussalam, Kalimantan Timur, dan DI Yogyakarta [2].

Berbagai upaya telah dilakukan oleh Pemerintah Indonesia melalui Badan Nasional Penanggulangan Bencana (BNPB) untuk menanggulangi masalah Covid-19, salah satunya dengan mengimpor alat kesehatan, alat kesehatan diagnostik *in vitro* dan perbekalan kesehatan rumah tangga yang digunakan untuk mitigasi [3].

Selain itu upaya yang telah dilakukan oleh pemerintah untuk mencegah transmisi penyakit yang lebih luas antara lain dengan menerapkan pembatasan sosial berskala besar dan memonitor wilayah-wilayah yang banyak terjadi migrasi penduduk lokal maupun asing. Provinsi Bali merupakan salah satu wilayah di Indonesia yang cukup banyak terjadi migrasi baik warga lokal maupun asing, mengingat Bali merupakan kawasan destinasi wisata yang populer. Walaupun Bali tidak termasuk dalam provinsi dengan kasus positif Covid-19 tertinggi di Indonesia, namun dengan maraknya wisatawan yang tetap berpergian ke Bali di tengah pandemi Covid-19 yang terjadi membuat masyarakat Bali maupun pemerintah provinsi tetap waspada terhadap penambahan kasus baru yang bersumber dari perjalanan masyarakat dari maupun ke Bali.

Penelitian ini bertujuan untuk melihat kejadian kasus positif Covid-19 berdasarkan jenis transmisi yang terjadi di seluruh wilayah Bali. Jika sumber penularan terbesar dapat diketahui sejak dini, maka pemerintah Provinsi Bali dapat mengambil tindakan tegas untuk menekan angka penambahan kasus positif Covid-19 dengan membatasi kegiatan atau perjalanan masyarakat baik yang akan masuk maupun keluar Bali. Oleh karena itu diharapkan penelitian ini dapat menjadi rekomendasi bagi pihak terkait untuk dapat merancang kebijakan atau strategi mitigasi yang sesuai dengan karakteristik penularan di setiap wilayah.

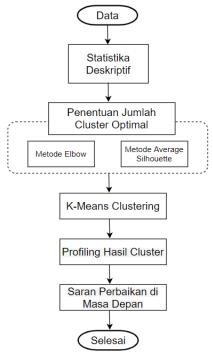
Metode yang digunakan untuk mengetahui karakteristik wilayah transmisi Covid-19 di Bali adalah *K-Means clustering*. *Clustering* bertujuan

untuk mengelompokkan objek ke dalam *cluster* di mana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya. *K-Means* adalah algoritma pengelompokan dalam bidang *data mining*. Metode ini digunakan untuk menganalisis *cluster* dan memiliki efisiensi tinggi pada pembagian data terutama dalam *dataset* besar [4].

II. METODOLOGI PENELITIAN

Data yang digunakan dalam penelitian ini adalah jumlah kasus positif Covid-19 pada setiap wilayah kabupaten/kota di Bali berdasarkan sumber penularan yang dilaporkan pada tanggal 19 Januari 2021 dari *website* Provinsi Bali Tanggap Covid-19 [5]. Sumber penularan yang dimaksud terdiri dari tiga jenis yaitu Riwayat Perjalanan Luar Negeri (PPLN), Riwayat Perjalanan Dalam Negeri (PPDN), dan Riwayat Transmisi Lokal (TL) [5]. Jumlah kasus positif di Bali berdasarkan ketiga jenis penularan tersebut sampai tanggal 19 Januari 2021 adalah sebanyak 21.929 kasus.

Metode yang digunakan dalam penelitian ini terdiri dari tiga tahap yaitu statistika deskriptif, analisis *cluster K-Means*, dan profilisasi hasil *cluster*. Alur penelitian ditunjukkan pada Gambar 1 berikut.



Gambar 1 Flowchart penelitian

A. Analisis Clustering

Analisis *cluster* membagi data menjadi kelompok (*cluster*) yang bermakna, berguna, atau keduanya. Jika kelompok yang bermakna adalah tujuan dari analisis *cluster*, maka *cluster* harus menangkap struktur alami data. Analisis *cluster* mengelompokkan objek data berdasarkan hanya pada informasi yang ditemukan dalam data yang mendeskripsikan objek dan hubungannya. Tujuannya adalah agar objek-objek dalam suatu kelompok menjadi serupa atau terkait satu sama lain dan berbeda dari atau tidak terkait dengan objek dalam kelompok lain [6].

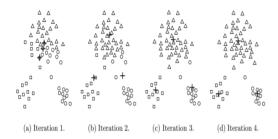
B. Algoritma K-Means

K-Means adalah teknik *cluster* partisional berbasis prototipe yang mencoba menemukan jumlah *cluster* (*K*) yang optimal, yang diwakili oleh sentroidnya. Langkah-langkah dalam algoritma *K-Means* dijelaskan oleh Tabel I [6].

TABEL I ALGORITMA *K-MEANS*

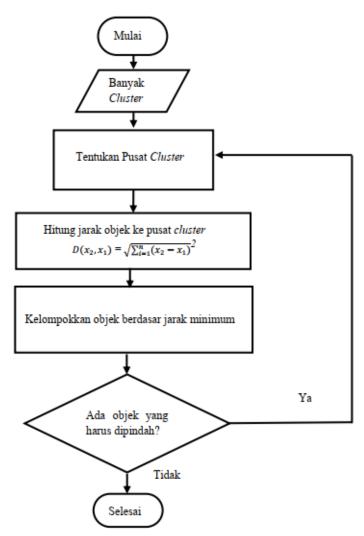
No	Langkah		
1	Pilih titik K sebagai sentroid awal		
2	Ulangi		
3	Bentuk <i>K cluster</i> dengan menetapkan setiap titik ke sentroid terdekatnya		
4	Hitung ulang sentroid dari setiap cluster		
5	sampai centroid tidak berubah		

Pada langkah pertama, ditunjukkan pada Gambar 2(a), titik ditetapkan ke sentroid awal, yang semuanya berada dalam kelompok titik yang lebih besar. Pada contoh ini, digunakan *mean* sebagai sentroid. Setelah titik ditetapkan ke sentroid, sentroid tersebut kemudian diperbarui. Gambar untuk setiap langkah menunjukkan sentroid pada awal langkah dan penetapan titik-titik ke sentroid tersebut. Pada langkah kedua, titik ditetapkan ke sentroid yang diperbarui dan sentroid diperbarui lagi.



Gambar 2 Menggunakan algoritma *K-Means* untuk menemukan tiga *cluster* dalam data sampel ([2])

Pada langkah 2, 3, dan 4 yang masing-masing ditunjukkan pada Gambar 2(b), (c), dan (d), dua sentroid bergerak ke kelompok kecil titik di bagian bawah gambar. Ketika algoritma *K-Means* berhenti dalam Gambar 2(d) karena tidak ada lagi perubahan yang terjadi, sentroid telah mengidentifikasi pengelompokan titik secara alami. Adapun tahapan dalam proses algoritma *K-Means* seperti ditunjukan pada Gambar 3.



Gambar 3 Flowchart algoritma K-Means ([7])

Keterangan:

- 1) Menentukan banyak *cluster* yang akan dibentuk.
- 2) Menentukan *K* sentroid (titik pusat *cluster*) secara acak.
- 3) Menghitung jarak setiap data terhadap masingmasing sentroid menggunakan rumus *Euclidean* berikut:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{1}$$

di mana:

 x_i adalah titik data *cluster* C_k μ_k adalah nilai rata-rata dari titik yang ditetapkan ke *cluster* C_k

4) Mengelompokkan data berdasarkan jarak terdekat antara data dengan sentroid.

 Menentukan nilai sentroid yang baru dengan menghitung rata-rata dari *cluster* yang bersangkutan menggunakan persamaan (2):

$$C_k = \frac{1}{n_k} \sum d_i \tag{2}$$

di mana:

 n_k adalah jumlah data dalam $cluster\ K$ d_i adalah jumlah nilai jarak yang masuk dalam masing-masing cluster

 Melakukan perulangan dari langkah 3–5 hingga anggota tiap *cluster* tidak ada yang berubah.

C. Jumlah Cluster Optimal

K-Means clustering mengharuskan pengguna untuk menentukan jumlah *cluster* yang akan dibuat. Berikut beberapa metode yang digunakan untuk menentukan jumlah klaster yang optimal [8]:

1. Metode Elbow

Seperti yang telah dijelaskan pada sub-bab sebelumnya bahwa *K-Means clustering* bertujuan untuk menentukan *cluster* sedemikian rupa sehingga variasi total dalam *cluster* diminimalkan:

$$minimize(\sum_{k=1}^{K} W(C_k))$$
 (3)

dimana C_k adalah *cluster* ke-k dan $W(C_k)$ adalah variasi dalam *cluster*. Jumlah kuadrat total dalam *cluster* (WSS) mengukur kekompakan dari pengelompokan dan diharapkan nilainya sekecil mungkin. Berikut merupakan algoritma untuk menentukan *cluster* yang optimal menggunakan metode *Elbow*:

- a) Menghitung algoritma *clustering* (misalnya: *K-Means clustering*) untuk nilai *K* yang berbeda. Misalnya dengan memvariasikan *K* dari 1 hingga 10 *cluster*.
- b) Untuk setiap *K*, hitung nilai WSS.
- c) Plot kurva WSS sesuai dengan jumlah *cluster K*.
- d) Lokasi belokan (*elbow*) di dalam plot umumnya dianggap sebagai indikator jumlah *cluster* yang sesuai.

2. Metode Average Silhouette

Metode *Average Silhouette* mengukur seberapa baik sebuah observasi dikelompokkan dan memperkirakan jarak rata-rata antar *cluster*. Plot *Silhouette* menampilkan ukuran seberapa dekat setiap titik dalam satu *cluster* dengan titik-titik di *cluster* tetangga. Untuk setiap pengamatan *i*, lebar siluet s_i dihitung sebagai berikut:

- a) Untuk setiap pengamatan i, hitung ratarata ketidaksamaan a_i antara i dan semua titik lain dari *cluster* tempat i berada.
- b) Untuk semua *cluster C* lainnya, yang bukan termasuk i, hitung perbedaan rata-rata d(i,C) dari i untuk semua pengamatan C. Rata-rata terkecil dari d(i,C) ini didefinisikan sebagai $b_i = \min_{C} d(i,C)$. Nilai b_i dapat dilihat sebagai ketidaksamaan antara i dan *cluster* tetangganya, yaitu *cluster* terdekat yang bukan miliknya.
- c) Lebar siluet dari pengamatan i didefinisikan dengan rumus:

$$S_i = \frac{(b_i - a_i)}{max(a_i, b_i)}$$
(4)

III. HASIL PENELITIAN

A. Statistika Deskriptif

digunakan Statistika deskriptif untuk meringkas data secara terorganisir dengan menggambarkan hubungan antar variabel dalam suatu sampel atau populasi. Menghitung statistik deskriptif merupakan langkah penting pertama saat melakukan penelitian dan harus selalu dilakukan membuat perbandingan inferensial [9]. Statistik deskriptif untuk variabel penelitian yang digunakan disajikan pada Tabel II. Sampel yang digunakan dalam penelitian ini adalah jumlah kasus positif Covid-19 pada setiap wilayah kabupaten/kota di Bali berdasarkan sumber penularan yang dilaporkan pada tanggal 19 Januari 2021 yang bersumber dari website Provinsi Bali Tanggap Covid-19 [4]. Jumlah kasus positif di Bali berdasarkan ketiga jenis penularan yaitu Riwayat Perjalanan Luar Negeri (PPLN), Riwayat Perjalanan Dalam Negeri (PPDN), dan Riwayat Transmisi Lokal (TL) sampai tanggal 19 Januari 2021 adalah sebanyak 21.929 kasus.

TABEL II STATISTIKA DESKRIPTIF UNTUK VARIABEL PENELITIAN

Ringkasan Data	PPDN	PPLN	TL
Rata-rata	45	28	1920
Nilai Minimum	3	6	34
Nilai Maksimum	211	58	5806

Berdasarkan Tabel II, dapat disimpulkan bahwa penularan atau variabel Perjalanan Luar Negeri (PPLN) menyumbangkan rata-rata kasus Covid-19 di Bali sebanyak 28 kasus. Kondisi yang hampir sama bersumber dari Riwayat Perjalanan Dalam Negeri (PPDN) dengan jumlah kasus hampir dua kali lipatnya. Sedangkan sumber Transmisi Lokal (TL) memberikan ratarata terbesar kasus positif Covid-19 di Bali mencapai 1920 kasus. Hal yang sama juga terlihat dari kasus terbanyak (maksimum) dari masingmasing sumber penularan, TL menjadi sumber penularan terbanyak disusul oleh PPDN dan PPDL.

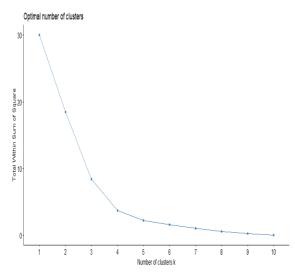
TABEL III JUMLAH KASUS POSITIF BERDASARKAN SUMBER PENULARAN PER KABUPATEN DI BALI

Kabupaten	Jumlah Kasus Positif Berdasarkan Sumber Penularan		
	PPDN	PPLN	TL
Badung	53	30	4060
Bangli	6	58	985
Buleleng	17	41	1473
Denpasar	211	48	5806
Gianyar	65	30	2629
Jembrana	34	26	1238
Kab. Lainnya	58	6	118
Karangasem	16	23	1126
Klungkung	3	19	1028
Tabanan	28	21	2624
Warga Negara	4	11	34
Asing			

Tabel III menunjukkan sebaran kasus positif Covid-19 di beberapa kabupaten di Bali berdasarkan sumber penularannya. Berdasarkan data yang dilaporkan pada website Provinsi Bali Tanggap Covid-19 kasus yang dilaporkan berdasarkan wilayah kabupaten di Bali dan juga Warga Negara Asing (WNA) yang tinggal di Bali. Oleh karena itu terdapat 9 kapupaten di Bali, 1 kabubaten lainnya, dan jumlah kasus positif khusus untuk WNA. Berdasarkan Tabel III terlihat bahwa kabupaten Denpasar memiliki jumlah kasus positif terbanyak yang bersumber dari transmisi lokal sebanyak 5.806 kasus. Sedangkan kasus positif terbanyak yang bersumber dari Riwayat Perjalanan Dalam Negeri (PPDN) dan Luar Negeri (PPLN) berturut-turut terjadi di kabupaten Denpasar sebanyak 211 kasus dan Bangli sebanyak 58 kasus.

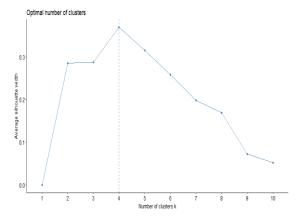
B. Analisis K-Means Clustering

Sebelum melakukan analisis *K-Means Clustering*, terlebih dahulu akan ditentukan jumlah *cluster* optimal dengan menggunakan kriteria WSS dan *Average Silhouette* [10].



Gambar 3 Metode WSS

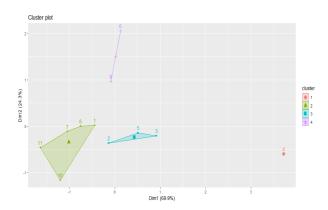
Metode Elbow menggunakan nilai total WSS sebagai penentu K optimalnya. Dari Gambar 4, garis mengalami patahan yang membentuk elbow atau siku pada saat K=4. Oleh karena itu dengan menggunakan metode ini diperoleh K optimal pada saat berada di K=4. Untuk menjadi pembanding, dilakukan uji Average Silhouette.



Gambar 4 Metode Average Silhouette

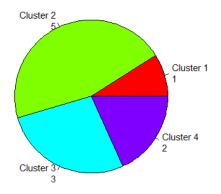
Berdasarkan Gambar 5, pendekatan rata-rata nilai metode *Average Silhouette* untuk menduga kualitas dari *cluster* yang terbentuk. Semakin tinggi nilai rata-ratanya maka akan semakin baik. Berdasarkan hasil Gambar 5 dipeoleh banyak *cluster* optimal yang terbentuk pada K = 4.

Gambar 6 berikut merupakan hasil plot *cluster* untuk kasus positif Covid-19 di beberapa kabupaten di Bali berdasarkan sumber penularan.



Gambar 5 Plot hasil K-Means Clustering

Gambar 6 menunjukkan hasil *K-Means Clustering* dengan jumlah *K* optimum sebanyak 4 berdasarkan hasil dari metode WSS dan *Average Silhouette*. Terlihat bahwa dari empat *cluster* yang terbentuk terjadi ketidakseimbangan jumlah anggota *cluster*. Secara sekilas, *cluster* dengan jumlah anggota yang terdapat perbedaan yang cukup menonjol adalah *cluster* 2 (berwarna hijau) dan *cluster* 1 (berwarna merah). Lebih lanjut, distribusi anggota *cluster* akan dijelaskan pada Gambar 7.



Gambar 7 Distribusi Hasil Cluster

Gambar 7 menunjukkan distribusi anggota dari masing-masing *cluster*. Pada *cluster* 1 beranggotakan hanya satu kabupaten yaitu Denpasar, *cluster* 2 terdiri dari lima kabupaten yaitu Jembrana, Klungkung, Karangasem, kabupaten lainnya, dan warga negara asing. Sedangkan kabupaten Tabanan, Badung, dan Gianyar masuk ke dalam satu *cluster* yang sama yaitu *cluster* 3. Sementara dua kabupaten sisanya yaitu Bangli dan Buleleng berada pada *cluster* 4.

IV. PEMBAHASAN

Proses *clustering* yang telah dilakukan menghasilkan empat *cluster* yang memiliki anggota masing-masing. Pada konsep *clustering*, setiap entitas dikelompokkan ke dalam *cluster* berdasarkan kemiripan atributnya, sehingga setiap *cluster* memiliki karakteristik atau profil yang membedakan antara *cluster* satu dengan yang lain. Pada Tabel IV berikut ditampilkan ringkasan karakteristik setiap *cluster* yang teridentifikasi.

 ${\it TABEL\,IV} \\ {\it RINGKASAN\,KARAKTERISTIK\,HASIL\,CLUSTER} \\$

Cluster	Rata-rata PPDN	Rata-rata PPLN	Rata-rata TL
1	211	48	5806
2	23	17	708,80
3	48,67	27	3140,33
4	11,50	49,50	1229

Tabel IV menunjukkan bahwa *cluster* 1 memiliki jumlah rata-rata kasus positif Covid-19 tertinggi berdasarkan ketiga sumber dibandingkan dengan *cluster* lainnya, sebaliknya *cluster* 2 menjadi *cluster* dengan kasus positif terendah baik yang bersumber dari PPDN, PPLN, maupun TL. Kasus positif tertinggi pada *cluster* 3 bersumber dari TL dan PPDN, sedangkan pada *cluster* 4 yang menjadi sumber penularan tertinggi yaitu dari TL dan PPLN.

Berdasarkan hasil *clustering* yang teridentifikasi, rekomendasi strategi mitigasi yang dapat diberikan kepada pihak terkait adalah dengan memperketat mobilitas penduduknya baik yang melakukan perjalanan dari dan ke luar negeri maupun dalam negeri. Terlebih menerapkan protokol kesehatan yang lebih ketat serta membatasi aktivitas masyarakat karena infeksi sudah terjadi di tengah masyarakat lokal (transmisi lokal). Penyebaran virus berdasarkan transmisi lokal inilah yang menjadi sumber penularan Covid-19 tertinggi di Bali.

V. KESIMPULAN

Penyebaran Covid-19 di Provinsi Bali secara rata-rata disebabkan oleh aktivitas transmisi lokal. Meskipun Bali menjadi destinasi wisata bagi warga negara asing, namun kasus positif yang disebabkan oleh perjalanan dalam negeri berjumlah lebih banyak dibanding kasus positif yang disebabkan oleh perjalanan luar negeri. Hasil penelitian menunjukkan terdapat empat *cluster* yang memiliki karakteristik yang berbeda, sehingga strategi mitigasi yang diusulkan perlu disesuaikan dengan karakteristik *cluster*. Mitigasi risiko infeksi untuk *cluster* ke-1, 2, dan 3 perlu difokuskan pada pembatasan transmisi lokal dan perjalanan dalam negeri. Sedangkan mitigasi untuk *cluster* ke-4 perlu difokuskan pada pembatasan transmisi lokal serta perjalanan luar negeri.

REFERENSI

- [1] CNN Indonesia, "55 Rumah Sakit Rujukan Covid-19 di Bali Hampir Penuh", CNN Indonesia, 9 September 2020. [Online]. Available: https://www.cnnindonesia.com/nasional/20200909150125-20-544480/55-rumah-sakit-rujukan-covid-19-di-balihampir-penuh/. [Använd 19 January 2021].
- [2] Kabar24, " 504 Tenaga Medis Wafat Akibat Virus Corona, Desember Paling Tinggi", Iim Fathimah Timorria, 2 January 2021. [Online]. Available: https://kabar24.bisnis.com/read/20210102/15/1337841/504tenaga-medis-wafat-akibat-virus-corona-desember-palingtinggi/. [Använd 19 January 2021].
- [3] Jakarta Globe, "Customs Office Simplifies Import Requirements for Medical Equipment", Mariska, D., 3 April 2020. [Online]. Available: https://jakartaglobe.id/news/customs-office-simplifiesimport-requirements-for-medical-equipment/. [Använd 19 January 2021].
- [4] Y. Yao, Y. Liu, Y. Yu, H. Xu, W. Lv, Z. Li, & X. Chen, "An Effective SVM Algorithm Based on K-Means Clustering," *Journal of Computers*, 8(10), 2013.
- [5] P. Bali, "Provinsi Bali Tanggap Covid-19," Provinsi Bali, 2 March 2020. [Online]. Available: https://infocorona.baliprov.go.id/. [Använd 19 January 2021].
- [6] P. N. Tan, M. Steinbach, A. Karpatne och V. Kumar, Introduction to Data Mining (Second Edition), New York CIty: Pearson, 2019.
- [7] F. A. Hermawati, Data Mining, Yogyakarta: Penerbit ANDI, 2013.
- [8] J. A. Hartigan och M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm," Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 28, nr 1, pp. 100-108, 1979.
- [9] L. Kaufman och P. J. Pousseeuw, Finding Groups in Data: An Introduction to *Cluster* Analysis, New Jersey: John Wiley & Sons, Inc., 1990.
- [10] P. Kaur, J. Stoltzfus och V. Yellapu, "Descriptive statistics," Int J Acad Med, pp. 60-63, 2018.