

IMPLEMENTASI *SPECTRAL CLUSTERING ALGORITHM* UNTUK PENGELOMPOKAN SASARAN VAKSINASI COVID-19 DI INDONESIA

Millenia Winadya Putri, *Indah Manfaati Nur, Rochdi Wasono
Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas
Muhammadiyah Semarang

e-mail : indahmnur@unimus.ac.id

ABSTRAK

Pandemi COVID-19 dalam kurun waktu dua tahun berhasil menginfeksi jutaan orang di seluruh dunia dan menyebabkan banyak kematian. Guna menghentikan penyebaran virus, pemerintah melakukan tindakan yaitu menerapkan protokol kesehatan dan mewajibkan vaksinasi kepada masyarakat. Namun, kegiatan vaksinasi masih lamban untuk mencapai target. Maka dari itu, dalam penelitian ini akan dilakukan suatu pengelompokan untuk mengetahui tingkat persebaran vaksinasi di Indonesia menurut provinsi dengan data jumlah vaksinasi per-kategori masyarakat pada tanggal 1 Februari 2022. Salah satu algoritma pengelompokan dalam Data Mining yaitu *Spectral Clustering*. Pengelompokan spektral merupakan teknik yang mengikuti pendekatan konektivitas, dimana metode ini mengklasifikasikan titik-titik yang terhubung atau berbatasan langsung. Penelitian ini menghasilkan 3 kluster untuk masing-masing kategori, yaitu kluster daerah-daerah yang memiliki tingkat persebaran vaksinasi tinggi, sedang, dan rendah. Evaluasi kluster diukur menggunakan Davies-Bouldin Index (DBI) dengan rata-rata nilai DBI tiap kategori yaitu 1,01422.

Keyword: COVID-19, Vaksinasi, Data Mining, *Spectral Clustering*, DBI

PENDAHULUAN

Coronavirus 2019 atau umum disebut COVID-19 merupakan jenis penyakit menular melalui udara yang pertama kali muncul pada Desember 2019 di Wuhan, China dan sedang menjadi pandemi global sejak awal tahun 2020 hingga saat ini. Virus ini menyebabkan penyakit infeksi saluran pernapasan, dari flu biasa sampai infeksi yang serius seperti *Middle East Respiratory Syndrome (MERS)* dan

Severe Acute Respiratory Syndrome (SARS). Kasus terkonfirmasi pada tahun 2020-2021, jutaan orang dari seluruh dunia jatuh sakit dan meninggal dunia di setiap harinya dan Indonesia menjadi salah satu negara yang paling terdampak pandemi COVID-19. Kondisi tersebut memberikan dampak langsung kepada masyarakat dunia, tidak terkecuali masyarakat Indonesia sehingga pemerintah Indonesia memberlakukan penerapan protokol kesehatan di seluruh

aspek kegiatan. Namun, protokol kesehatan dinilai kurang mampu dalam mencegah risiko penularan COVID-19. Oleh karena itu, pemerintah perlu melakukan tindakan intervensi lain yang efektif guna memutus rantai penyebaran virus yaitu melalui upaya vaksinasi. Tujuan utama vaksinasi yaitu untuk menurunkan angka kesakitan dan kematian yang disebabkan oleh COVID-19, mencapai kekebalan dan melindungi masyarakat dari COVID-19, kemudian puncaknya dapat menjaga masyarakat dan perekonomian (Kemenkes RI Dirjen P2P, 2020).

Vaksinasi diberlakukan kepada masyarakat dengan mengacu pada kelompok yang telah ditetapkan oleh pemerintah agar vaksin yang akan diberikan merupakan vaksin yang sesuai dengan kondisi atau keadaan penerimanya. Kelompok masyarakat tersebut dibagi menjadi lima, yaitu Lansia, Masyarakat Umum dan Rentan, Petugas Publik, SDM Kesehatan, serta Usia 12-17 Tahun (Direktorat Jenderal Pencegahan dan Pengendalian Penyakit (P2P), 2020). Saat ini, vaksinasi belum juga dapat mencapai target sasaran. Karenanya pemerataan kegiatan vaksinasi di seluruh wilayah Indonesia harus ditinjau lebih dalam lagi untuk mengupayakan terpenuhinya target sasaran vaksinasi COVID-19 nasional. Upaya tersebut dapat dilakukan dengan melakukan suatu penelitian untuk meninjau persebaran vaksinasi yaitu dengan mengelompokkan wilayah mana saja yang kegiatan vaksinasinya masih rendah, dilihat dari total masyarakat di wilayah tersebut yang sudah mendapatkan vaksin dosis pertama. Belum sepenuhnya pelaksanaan vaksinasi dosis pertama menjadi

pertimbangan penggunaan datanya dalam penelitian. Pengelompokan didasarkan pada setiap kelompok masyarakat penerima vaksin.

Penelitian ini sangat penting dilakukan karena dampak negatif yang diakibatkan oleh pandemi ini tidak akan terselesaikan dengan cepat jika kasus terkonfirmasi positif COVID-19 masih tinggi. Dampak yang dihasilkan tidak hanya pada sektor kesehatan saja, namun dampak serius terjadi pada sektor lain, seperti sektor sosial, sektor ekonomi, sektor industri, serta sektor pariwisata yang tidak dapat beroperasi secara maksimal. Sehingga, dengan adanya penelitian ini diharapkan mampu mengetahui daerah-daerah yang masih sedikit tingkat vaksinasinya agar dapat dilakukan tindakan pendistribusian vaksin dan target sasaran vaksinasi segera terpenuhi.

METODE PENELITIAN

Sumber Data dan Variabel Penelitian

Jenis penelitian ini adalah kuantitatif dengan desain *cross sectional* dimana populasi penelitian adalah seluruh masyarakat Indonesia dengan sampel kelompok masyarakat penerima vaksin, yaitu Lansia, Masyarakat Umum dan Rentan, Petugas Publik, SDM Kesehatan, serta Usia 12-17 Tahun. Data yang digunakan merupakan data jumlah vaksinasi COVID-19 per-Provinsi tiap kelompok masyarakat pada tanggal 1 Februari 2022. Data tersebut adalah data sekunder yang didapatkan dari laman *website* vaksin.kemkes.go.id. Penelitian ini dilakukan menggunakan beberapa tahapan, yaitu pengumpulan data, analisis deskriptif dengan *descriptive statistics*, *data pre-processing*, pengelompokan

data menggunakan algoritma *Spectral Clustering*, serta evaluasi kluster dengan nilai *Davies-Bouldin Index* (DBI).

Metode Analisis

Langkah analisis yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Menyiapkan data akan dianalisis.
2. Mendeskripsikan data dengan analisis deskriptif untuk mengetahui gambaran umum kegiatan vaksinasi COVID-19 menurut provinsi di Indonesia.
3. Membentuk Matriks Similaritas.
4. Membentuk Matriks Diagonal.
5. Membentuk Matriks Laplacian.
6. Menghitung eigenvalue dan eigenvector.
7. Mengelompokkan data dengan *K-Means Clustering*.
8. Menginterpretasi hasil kluster berdasarkan karakteristik tiap kluster.
9. Mengevaluasi hasil kluster dengan nilai DBI.

HASIL DAN PEMBAHASAN

Analisis Deskriptif

Tabel 1. Hasil Statistika Deskriptif

	Lansia	Masyarakat Umum & Rentan	Petugas Publik	SDM Kesehatan	12-17 Tahun
Min	7198	152832	73699	8922	22487
Mean	278166	2190584	829366	59539	4379170
Median	804134	927556	334921	35063	206549
Max	1954096	14353940	4809415	275413	2532656

Berdasarkan tabel, informasi yang dapat diambil yaitu provinsi dengan pelaksanaan vaksinasi paling rendah, antara lain provinsi papua barat untuk kategori lansia, kategori masyarakat umum dan rentan, serta kategori usia 12-17 tahun, provinsi kalimantan utara untuk kategori petugas publik dan kategori sdm kesehatan. Sedangkan provinsi dengan

pelaksanaan vaksinasi paling tinggi antara lain, provinsi jawa tengah untuk kategori lansia, provinsi jawa barat untuk kategori masyarakat umum dan rentan, kategori petugas publik, serta kategori usia 12-17 tahun, dan provinsi jawa timur untuk kategori sdm kesehatan.

Data Pre-processing

Tabel 2. Hasil Standarisasi Data

	Lansia	Masyarakat Umum & Rentan	Petugas Publik	Sdm Kesehatan	Usia 12-17 Tahun
1	0.12506198	-0.2306964	-0.02490849	-0.125060601	-0.43404743
2	0.26454564	0.273595	0.45284048	0.517404394	0.91081771
3	-0.19315414	-0.1717328	-0.09322556	0.008958858	0.07211049
...
32	0.05335381	0.4525591	0.25640544	0.137151384	0.22662308
33	0.0403646	-0.1779333	0.25316004	-0.098607729	-0.09554145
34	-0.45400451	-0.4030999	-0.42284204	0.071370225	-0.40153314

Tahap preprocessing dilakukan untuk mempersiapkan data sehingga siap digunakan. Pada penelitian ini, *preprocessing* yang dilakukan yaitu *scaling* data dengan metode standarisasi.

Clustering Data

Proses pengelompokan data dilakukan dengan algoritma *Spectral Clustering*, dimana terdapat beberapa tahap perhitungan matriks didalamnya. Berikut tahapan pada *Spectral Clustering*:

Membentuk Matriks Similaritas

Matriks similaritas dibentuk menggunakan metode *k-nearest neighbors* dimana prinsip kerjanya yaitu mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga (*neighbors*) terdekatnya yang dihitung dengan rumus *euclidean distance* (yudhana dkk, 2020). Pada penelitian ini, inisialisasi nilai k menggunakan nilai 10 dengan pertimbangan nilai k yang optimum adalah tidak terlalu kecil dan tidak terlalu besar, serta mempertimbangkan jumlah data. Hasil *similarity matrix* direpresentasikan dengan angka 0 dan 1 dimana 0 berarti antar data tidak berdekatan dan 1 berarti

antar data saling berdekatan yang dapat dilihat pada tabel berikut:

Tabel 3. Matriks Similaritas

	V1	V2	V3	...	V32	V33	V34
1	0	0	1	...	0	1	0
2	1	0	1	...	1	1	0
3	1	0	0	...	0	1	0
...
32	1	1	1	...	0	1	0
33	1	0	1	...	0	0	0
34	0	0	0	...	0	0	0

Membentuk Matriks Diagonal

Matriks derajat atau matriks diagonal ini merepresentasikan jumlah data yang saling terhubung atau berdekatan berdasarkan matriks similaritas (Ramdan & Hermawan, 2012). Matriks diagonal merupakan matriks identitas dimana seluruh elemen bernilai 0 kecuali pada diagonal utamanya. Hasil matriks diagonal ditunjukkan pada tabel berikut:

Tabel 4. Matriks Diagonal

	V1	V2	V3	...	V32	V33	V34
1	9	0	0	...	0	0	0
2	0	6	0	...	0	0	0
3	0	0	11	...	0	0	0
...
32	0	0	0	...	6	0	0
33	0	0	0	...	0	11	0
34	0	0	0	...	0	0	10

Membentuk Matriks Laplacian

Matriks *laplacian* merupakan matriks representasi dari sebuah graf. Matriks *laplacian* ini sebagai komponen utama yang nantinya akan digunakan untuk mengelompokkan data dengan *spectral clustering*. Berikut adalah tabel matriks laplacian yang dihitung berdasarkan rumus yang ternormalisasi atau *normalized laplacian*:

Tabel 5. Matriks Laplacian

	V1	V2	V3	...	V32	V33	V34
1	0	0	0.003086420	...	0	0.003086420	0
2	0.006944444	0	0.006944444	...	0.006944444	0.006944444	0
3	0.002066116	0	0	...	0	0.002066116	0
...
32	0.006944444	0.006944444	0.006944444	...	0	0.006944444	0
33	0.002066116	0	0.002066116	...	0	0	0
34	0	0	0	...	0	0	0

Menghitung nilai eigen dan vektor eigen

Eigenvalue dan *eigenvector* berguna untuk menghasilkan data baru yang merepresentasikan karakteristik kesamaan antar data dengan lebih dalam (Rohmah & Sugiyarto, 2021). Pertama, dilakukan penghitungan nilai *eigenvalue* dan setelah diketahui maka selanjutnya dibentuk *eigenvector* berdasarkan tiga *eigenvalue*. Pada tahap ini, pembentukan *eigenvector* dilakukan untuk setiap kategori agar terbentuk *cluster* pada setiap kategori. Berikut adalah salah satu tabel *eigenvector* yang telah terbentuk:

Tabel 6. Vektor Eigen Kategori Lansia

Lansia		
V1	V2	V3
0.015459	0.149946	-0.176856
0.041982	0.172369	-0.176125
...
0.015459	0.149946	-0.176856
0.015459	0.149946	-0.176856
0.159922	0.024471	0.225777

Tabel 7. Vektor Eigen Kategori Masyarakat Umum dan Rentan

Masyarakat Umum dan Rentan		
V1	V2	V3
-0.044850	0.132246	0.170200
-0.007392	0.157138	0.169071
...
-0.007392	0.157138	0.169071
-0.007392	0.157138	0.169071
0.089633	-0.271547	-0.012696

Tabel 8. Vektor Eigen Kategori Petugas Publik

Petugas Publik		
V1	V2	V3
-0.044850	0.132246	0.170200
-0.007392	0.157138	0.169071
...
-0.007392	0.157138	0.169071
-0.007392	0.157138	0.169071
0.089633	-0.271547	-0.012696

Tabel 9. Vektor Eigen Kategori SDM Kesehatan

SDM Kesehatan		
V1	V2	V3
-0.036878	0.041504	-0.175717
0.020902	0.130311	-0.179792
...
0.020902	0.130311	-0.179792
-0.031818	0.073207	-0.179155
-0.021582	0.105289	-0.180983

Tabel 10. Vektor Eigen 12-17 Tahun

Usia 12-17 Tahun		
V1	V2	V3
-0.110434	0.271789	-0.018881
-0.066353	-0.166045	0.170886
...
-0.008237	-0.144110	0.172721
0.060488	-0.087296	0.172536
-0.029839	0.292290	0.014116

Mengelompokkan Data dengan K-Means

Hasil *eigenvector* setiap kategori selanjutnya dikelompokkan menggunakan *k-means*. Hasil *cluster* diinterpretasikan berdasarkan nilai rata-rata dari data vaksinasi covid-19. Hasil *cluster* persebaran vaksinasi tinggi, sedang, dan rendah ditunjukkan pada tabel di bawah ini:

Tabel 11. Kluster Kategori Lansia

Cluster	Rata-Rata	Anggota
1	742051,2	Yogyakarta, Sumatera Utara, Sumatera Selatan, NTB, Lampung, Jawa Timur, Jawa Tengah, Jawa Barat, DKI Jakarta, Banten, Bali
2	64841,93	Sumatera Barat, Sulawesi Tengah, Riau, NTT, Maluku, Kep. Riau, Kep. Bangka Belitung, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat, Jambi, Bengkulu, Aceh
3	43004,67	Sulawesi Utara, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Barat, Papua Barat, Papua, Maluku Utara, Kalimantan Utara, Gorontalo

Tabel 12. Kluster Kategori Masyarakat Umum dan Rentan

Cluster	Rata-Rata	Anggota
1	309742,0909	Sulawesi Tenggara, Sulawesi Tengah, Sulawesi Barat, Papua Barat, Papua, Maluku Utara, Maluku, Kep. Bangka Belitung, Kalimantan Utara, Gorontalo, Bengkulu
2	922800,3333	Yogyakarta, Sumatera Barat, Sulawesi Utara, Riau, NTT, Kep. Riau, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat, Jambi, Aceh
3	5453198,182	Sumatera Utara, Sumatera Selatan, Sulawesi Selatan, NTB, Lampung, Jawa Timur, Jawa Tengah, Jawa Barat, DKI Jakarta, Banten, Bali

Tabel 13. Kluster Kategori Petugas Publik

Cluster	Rata-Rata	Anggota
1	2006020,091	Yogyakarta, Sumatera Utara, Sumatera Selatan, Sulawesi Selatan, Riau, Jawa Timur, Jawa Tengah, Jawa Barat, DKI Jakarta, Banten, Bali
2	266373,1	Sumatera Barat, Sulawesi Utara, Sulawesi Tenggara, Sulawesi Tengah, Papua, NTB, Kep. Bangka Belitung, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Barat, Aceh
3	651699,9167	Sulawesi Barat, Papua Barat, NTT, Maluku Utara, Maluku, Lampung, Kep. Riau, Kalimantan Utara, Kalimantan Timur, Jambi, Gorontalo, Bengkulu

Tabel 14. Kluster Kategori SDM Kesehatan

Cluster	Rata-Rata	Anggota
1	17417,21429	Sulawesi Utara, Sulawesi Tenggara, Sulawesi Tengah, Sulawesi Barat, Papua Barat, Papua, Maluku Utara, Maluku, Kep. Riau, Kep. Bangka Belitung, Kalimantan Utara, Kalimantan Tengah, Gorontalo, Bengkulu
2	38975,44444	Sumatera Barat, Riau, NTT, NTB, Lampung, Kalimantan Timur, Kalimantan Selatan, Kalimantan Barat, Jambi
3	129970	Yogyakarta, Sumatera Utara, Sumatera Selatan, Sulawesi Selatan, Jawa Timur, Jawa Tengah, Jawa Barat, DKI Jakarta, Banten, Bali, Aceh

Tabel 15. Klaster Kategori Usia 12-17 Tahun

Cluster	Rata-Rata	Anggota
1	777650	Yogyakarta, Sulawesi Utara, Sulawesi Tengah, Kep. Riau, Kep. Bangka Belitung, Kalimantan Tengah, Jawa Timur, Jawa Tengah, Jawa Barat, Bengkulu
2	419367,0625	Sumatera Utara, Sumatera Selatan, Sumatera Barat, Sulawesi Selatan, Riau, NTT, NTB, Lampung, Kalimantan Timur, Kalimantan Selatan, Kalimantan Barat, Jambi, DKI Jakarta, Banten, Bali, Aceh
3	50290,875	Sulawesi Tenggara, Sulawesi Barat, Papua Barat, Papua, Maluku Utara, Maluku, Kalimantan Utara, Gorontalo

Evaluasi Data

Tahap evaluasi dilakukan menggunakan *davies-bouldin index* (dbi) untuk menentukan hasil *cluster* paling optimal dari proses *clustering*. Penentuan *cluster* optimal dapat disimpulkan melalui nilai dbi yang terendah atau dapat dikatakan bahwa semakin rendah nilai dbi, maka semakin baik hasil *cluster* yang didapatkan. Berikut adalah tabel nilai dbi masing-masing kategori:

Tabel 16. Tabel Nilai DBI

Kategori	DBI
Lansia	0,9910664
Masyarakat Umum & Rentan	1,022632
Petugas Publik	1,039109
SDM Kesehatan	0,9899738
Usia 12-17 Tahun	1,028337

KESIMPULAN DAN SARAN

Berdasarkan hasil dan diskusi penelitian ini, dapat disimpulkan bahwa setiap kategori masyarakat dikelompokkan menjadi tiga klaster, yaitu klaster dengan tingkat vaksinasi tinggi, sedang, dan rendah. Evaluasi hasil klaster menunjukkan nilai yang hampir sama sehingga dapat terlihat bahwa anggota-anggota pada suatu klaster di masing-masing kategori tidak jauh berbeda. Titik fokus berada pada tingkat vaksinasi rendah karena nantinya klaster tersebut yang akan diperhatikan untuk proses pendistribusian vaksin agar target sasaran vaksinasi cepat tercapai.

DAFTAR PUSTAKA

- [1] Direktorat Jenderal Pencegahan dan Pengendalian Penyakit (P2P). (2020). "Pedoman pencegahan dan pengendalian coronavirus disease (covid-19)". Jakarta: Kementerian Kesehatan RI.
- [2] Kemenkes RI Dirjen P2P. (2020). "Kementerian Kesehatan Republik Indonesia, Tentang Petunjuk Teknis Pelaksanaan Vaksinasi Dalam Rangka Penanggulangan Pandemi Corona Virus Disease 2019 (Covid-19)". Kementerian Kesehatan RI, 5(1), 1.
- [3] Ramdan, D., dan Hermawan, G. (2012). "Pemanfaatan Metode Adjacency Matrix untuk Optimasi Rute Jalan Berbasis Web". Jurnal Komputer dan Informatika, 1(1), 59-63.
- [4] Rohmah, Q., dan Sugiyarto. (2021). "Implementasi Algoritma Spectral Clustering untuk Analisis Sentimen". Jurnal Ilmiah Pendidikan Matematika, 9(1), 27-36.
- [5] Yudhana, A., Sunardi., dan Hartanta, A. J. S. (2020). "Algoritma K-NN dengan Euclidean Distance untuk Prediksi Hasil Penggajian Kayu Sengon". Transmisi 22(4): 123-129.