

PERBANDINGAN TEKNIK CLUSTERING ANTARA METODE K-MEANS, K-MEDOIDS, DAN GSDMM PADA DATA TWEET VAKSINASI INDONESIA

Vauwez Sam El Fareez¹; Rofiqah Julia Nurlaila Saing²; Ainul Haqi³;

2009106054¹; 1915016013²; 1915016016³

Informatika - Fakultas Teknik - Universitas Mulawarman



ABSTRAK

Kegiatan vaksinasi merupakan salah satu langkah yang diambil pemerintah untuk menekan angka laju penambahan pasien Covid-19. Dengan adanya kegiatan vaksinasi tersebut tentunya mengundang berbagai tanggapan dan pembahasan masyarakat utamanya di media sosial mengenai kegiatan vaksinasi tersebut. Salah satu media sosial yang banyak digunakan adalah twitter. Banyaknya tanggapan dan topik pembahasan masyarakat mengenai kegiatan vaksinasi ini merupakan peluang data yang dapat dilakukan penambangan. Salah satu Teknik dalam melakukan penambangan data ini yaitu Teknik *clustering*. Dalam melakukan *clustering* terdapat beberapa metode. Diantaranya yang akan dilakukan dalam penelitian ini yaitu metode K-means, K-medoids, dan GSDMM. Pada penelitian ini digunakan sebanyak 10 cluster untuk diterapkan pada metode-metode tersebut. Tujuan dan kontribusi utama dari penelitian ini adalah untuk melakukan analisa dan perbandingan dari ketiga metode *Clustering* yaitu K-means, K-medoids, dan GSDMM. Dari hasil *labelling* tiap metode didapatkan kecenderungan pembahasan masyarakat mengenai vaksinasi adalah tentang jenis vaksin. Selain itu, dilakukan pula perbandingan tiap metodenya. Hasil perbandingan tiap metode berdasarkan *silhouette index*, *davies-bouldin* dan waktu eksekusi didapatkan metode k-means adalah sebagai metode terbaik.

Kata kunci: *K-means*, *K-medoids*, *GSDMM*, *Clustering*, *Silhouette Index*, *Davies-Bouldin*.

1. PENDAHULUAN

1.1 Latar Belakang

Terhitung tanggal 2 Maret 2020, untuk pertama kalinya pemerintah mengumumkan dua kasus pasien positif Covid-19 di Indonesia. Sebuah wabah virus yang berasal dari Wuhan, China yang sudah menelan banyak korban di hampir seluruh dunia hingga ditetapkan sebagai pandemik global oleh World Health Organization (WHO) pada tanggal 11 Maret 2020 lalu. Berawal dari dua kasus pasien positif tersebut, jumlah kasus yang menjangkit masyarakat Indonesia terus bertambah setiap harinya. Hingga terus mencatat rekor pertambahan kasus positif dalam sehari seperti pada tanggal 14 Juli 2021 yang tercatat sebanyak 56.757 pasien baru dalam sehari. Berbagai upaya telah dilakukan untuk menanggapi kondisi yang memberikan dampak langsung kepada jutaan bahkan seluruh

masyarakat dunia ini. Mulai dari diberlakukannya protokol Kesehatan yang harus ditetapkan pada seluruh aspek kegiatan, pembatasan sosial hingga *lockdown* total sehingga menghambat seluruh kegiatan masyarakat.

Melihat kondisi penyebaran kasus Covid-19 yang mulai tidak terkendali dan mulai memunculkan berbagai mutasi baru seperti varian alpha, beta, dan delta, pada tanggal 14 Januari 2021 pemerintah mulai melakukan kegiatan vaksinasi di Indonesia. Hingga per Rabu (21/7/2021), menurut data satgas Covid-19 total jumlah vaksinasi pertama sudah mencapai 42.611.602. Adapun total jumlah penerima vaksinasi kedua di Indonesia mencapai 16.606.675 dan akan terus digencarkan untuk berbagai daerah dengan target vaksinasi mencapai 181 juta agar dapat menimbulkan imunitas kelompok atau *herd immunity*.

Dengan adanya kegiatan vaksinasi tersebut tentunya mengundang berbagai tanggapan dan pembahasan masyarakat utamanya di media sosial mengenai kegiatan vaksinasi tersebut. Berdasarkan data dari Global Digital Statistic “Digital, Social & Mobile in 2019” di We Are Social (2019), pada tahun 2019 jumlah pengguna media sosial di Indonesia yaitu berjumlah lebih dari 150 juta pengguna. Salah satu media sosial yang sering digunakan masyarakat untuk menyampaikan tanggapan atau pendapat adalah twitter. Media sosial ini dianggap sebagai media yang mudah untuk digunakan dan sangat cepat dalam menyebarkan informasi. Di Indonesia sendiri, penggunaannya mencakup lebih dari 52 persen dari total pengguna media sosial di Indonesia. Terkait dengan kegiatan vaksinasi, tentunya tidak sedikit masyarakat Indonesia yang menyampaikan tanggapan dan pendapatnya di media sosial twitter ini.

Hal tersebut menunjukkan adanya peluang sumber data yang sangat besar yang dapat diolah atau dilakukan penggalian data yang dalam penelitian ini disebut data mining untuk menghasilkan suatu *knowledge* yang bermanfaat. Beberapa penelitian telah dilakukan terkait topik vaksinasi di Indonesia. Salah satunya yaitu penelitian yang dilakukan oleh (Fajar Fathur Rachman & Setia Pramana, 2020) dengan judul Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin Covid-19 pada Media Sosial Twitter. Pada penelitian tersebut peneliti menggunakan metode Latent Dirichlet Allocation (LDA) untuk mengklasifikasikan respon masyarakat ke dalam sentiment positif & negative, dengan data yang digunakan adalah data tweets yang diposting di media sosial twitter pada rentang tanggal 25 oktober- 3 november 2020. Hasil yang didapat dari penelitian tersebut menghasilkan jumlah tweets 1461 tweets untuk sentiment positif, 2313 tweets untuk sentiment netral dan 1167 tweets untuk sentiment negative.

Adapun pada penelitian kali ini akan dilakukan salah satu Teknik dalam data mining yaitu Teknik *clustering* untuk memetakan atau melihat apa saja topik atau bahasan masyarakat yang terkait dengan vaksinasi. *Clustering* yang digunakan pada penelitian ini menggunakan metode GSDMM (*Gibbs Sampling algorithm for Dirichlet Multinomial Mixture*), K-Means, dan K-Medoids. Hasil dari metode-metode tersebut akan dibandingkan berdasarkan beberapa kriteria atau karakteristik untuk mengetahui perbedaan performa masing-masing metode untuk data tweet vaksinasi Indonesia pada penelitian ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas bahwa masalah yang dihadapi saat ini yaitu bagaimana menerapkan suatu metode data mining *Clustering* dalam mengelola data tweet vaksinasi Indonesia. Adapun hasil latar belakang di atas dapat dirumuskan sebagai berikut:

1. Bagaimana mempelajari dan menganalisa ketiga metode clustering yaitu K-means, K-medoids, dan GSDMM?
2. Bagaimana membandingkan hasil dari penerapan Teknik *clustering* dengan menggunakan metode K-means, K-medoids, dan GSDMM?
3. Bagaimana menentukan metode clustering terbaik dengan menggunakan 3 aspek yaitu *Silhouette index*, waktu eksekusi dan *Davies-Bouldin Index*.

1.3 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini, yaitu:

1. Untuk mengetahui penerapan dan hasil dari Teknik *clustering* dengan menggunakan metode K-Means, K-Medoids, dan GSDMM.
2. Dapat mengetahui perbedaan performa, karakteristik dan kriteria masing-masing metode pada data tweet vaksinasi Indonesia.
3. Dapat mengetahui metode terbaik dalam melakukan *clustering* pada data tweet vaksinasi Indonesia.

1.4 Manfaat

Penggunaan beberapa metode untuk Teknik *clustering* pada penelitian ini, diharapkan dapat menambah referensi baru untuk peneliti di bidang data mining dalam membandingkan dan memilih metode *clustering* untuk data yang diolah khususnya pada data korpus teks Bahasa Indonesia.

Dengan hasil *labelling* dari Teknik *clustering* yang dilakukan yaitu menghasilkan gambaran besar dari tanggapan atau topik pembahasan masyarakat mengenai vaksinasi di Indonesia, diharapkan bermanfaat bagi tenaga Kesehatan atau pemerintah dalam memberikan kebijakan dalam kegiatan vaksinasi kedepannya.

1.5 Batasan Masalah

Mengingat luasnya ruang lingkup penelitian dalam implemantasi teknik-teknik data mining, maka penelitian ini dibatasi pada:

1. Data yang digunakan dalam penelitian ini hanya data tweet vaksinasi Indonesia yang didapatkan dengan metode web scraping menggunakan bantuan website TAGS (*Twitter Archiving Google Sheet*) pada tanggal 27 Juli 2021.
2. Teknik data mining yang digunakan adalah Teknik *Clustering*.
3. Metode yang digunakan hanya tiga yaitu GSDMM, K-Means, dan K-Medoids.
4. Dalam membandingkan metode, hanya menggunakan kriteria yang telah tertera dalam penelitian ini.
5. Penambahan data dalam penelitian ini menggunakan Bahasa pemrograman python.

2. TINJAUAN STUDI

2.1 Clustering

Salah satu Teknik dalam melakukan penambahan data adalah dengan *clustering*. Teknik *clustering* ini merupakan penambahan data dengan melakukan proses pengelompokan data dengan mempertimbangkan pola-pola dalam suatu cluster agar mendapatkan hasil *cluster* dengan tingkat kemiripan tinggi dalam satu *cluster* dan tingkat kemiripan rendah dengan *cluster* lain. Hasil Teknik clustering ini bermanfaat dalam membuat keputusan dengan melihat pola-pola data dan kelompok datanya. Ada beberapa metode dalam clustering, diantaranya yaitu k-means, k-medoids, dan GSDMM.

2.1.1. K-Means Cluster

Metode K-means membentuk *cluster* dengan mengatur data ke dalam nilai K sebagai jumlah *cluster* yang telah ditetapkan. Kemudian, akan dilakukan perhitungan jarak data terhadap masing-masing titik pusat (*centroid*) yang telah ditetapkan secara acak. Objek akan dikelompokkan berdasarkan jarak terendah. Lalu, *centroid* baru akan ditentukan dengan menggunakan sebuah persamaan yang menghitung rata-rata dari objek yang ada di dalam *cluster*. Iterasi akan berulang seterusnya hingga tidak terjadi perpindahan anggota *cluster* ke *cluster* lain. Objek data yang memiliki pola atau karakteristik yang sama akan dimasukkan ke dalam satu *cluster*. Sedangkan objek data dengan pola atau karakteristik yang berbeda akan berada pada *cluster* lain.

2.1.2. K-Medoids Cluster

Jika pada metode k-means penentuan titik pusat (*centroid*) dilakukan dengan menghitung rata-rata, pada metode k-medoids ini dilakukan perhitungan titik acuan (*medoid*) dengan memilih secara acak k objek dari sebuah data. Tahapan-tahapan pada metode k-medoids adalah:

- K-medoids merupakan metode dalam clustering yang memilih k objek secara acak dalam sekumpulan data sebagai medoid (titik acuan)
- Pengelompokan objek dilakukan berdasarkan jarak terdekat dari setiap objek yang bukan titik acuan
- Untuk medoid baru dipilih secara acak dari objek yang bukan merupakan medoid sebelumnya
- Kemudian, dari pertukaran medoid, dihitung total cost S. nilai S yang lebih kecil dari nol ditukar objek secara random agar dapat membentuk sekumpulan k objek baru sebagai titik acuan. Iterasi akan berulang hingga tidak ada perubahan posisi.

2.1.3. GSDMM Cluster

GSDMM (*Gibbs Sampling Dirichlet Multinomial Mixture*) adalah metode *clustering* teks pendek. Model ini diklaim dapat mengatasi masalah pada teks pendek sementara juga menampilkan topik seperti model LDA. GSDMM pada dasarnya adalah LDA yang dimodifikasi yang mengasumsikan bahwa dokumen (tweet) mencakup hanya 1 topik.

Prinsip GSDMM dideskripsikan menggunakan analogi "Pendekatan Grup Film". Misalnya ada kelompok siswa (dalam hal ini adalah dokumen) yang masing masing punya list film favoritnya (dalam hal ini adalah kata). Tujuannya adalah membuat *cluster* siswa yang mana setiap grup adalah siswa yang punya minat film yang sama. Siswa-siswa tadi secara acak ditaruh di K meja. Satu per satu, setiap siswa harus berpindah ke meja yang lain dengan tujuan:

1. Cari meja dengan siswa yang lebih banyak.
2. Pilih meja dimana siswanya punya minat film yang sama.

Setelah mengulang-ulang proses ini, diekspektasikan ada meja yang hilang (tak beranggota) dan meja lain semakin banyak anggotanya dan akhirnya terdapat *cluster* siswa yang minat filmnya sama.

3. METODOLOGI PENELITIAN

3.1 Data

Untuk mengambil data tanggapan atau topik pembahasan masyarakat mengenai kegiatan vaksinasi Covid-19 di Indonesia pada media sosial twitter, dilakukan Teknik web scraping. Kegiatan web scraping menggunakan bantuan website TAGS (*Twitter Archiving Google Sheet*) yang di lakukan pada tanggal 27 Juli 2021 jam 17.20 WITA.

Pada penelitian ini “vaksin” dan “vaksinasi” menjadi kata kunci untuk melihat dan mengambil data postingan tweets masyarakat mengenai vaksinasi di Indonesia. Karena adanya keterbatasan pengumpulan data, penelitian ini menggunakan data tweets 8 hari terakhir yang diposting di media sosial twitter dengan rentang tanggal 20-27 Juli 2021.

3.2 Metode Persiapan Data

Data yang diperoleh dengan web *scraping*, tidak dapat langsung dilakukan penambangan data, akan tetapi perlu melalui tahapan *preprocessing data*. Dimana data yang

akan diolah yang dalam penelitian ini adalah data teks Bahasa Indonesia akan melalui beberapa tahap, yaitu:

- *Mention* dihilangkan dan semua teks diubah menjadi *lowercase*
- *Url link* dihilangkan
- *Hashtag* dihilangkan
- Angka dihilangkan
- Semua kecuali kata dan spasi dihilangkan
- Mengganti kata-kata singkatan (*contractions*) (dalam Bahasa Inggris jika ada)
- Kembalikan setiap kata ke kata dasar
- Mengganti singkatan (kontraksi) Bahasa Indonesia
- Hilangkan *stopwords* Bahasa Indonesia
- Hapus spasi yang berlebihan

3.3 Metode Analisis Data

Data yang telah dilakukan pembersihan, akan dibuat term-document matrix. Pembobotan TF dan TF-IDF akan dilakukan setelah baris dari setiap matrix berisi term dan setiap kolomnya untuk dokumen.

Kemudian, dalam melakukan penambangan data pada data tweet vaksinasi Indonesia, pada penelitian menggunakan Teknik *clustering* dengan beberapa metode yaitu k-means, k-medoids, dan GSDMM. Pada metode-metode tersebut akan dilakukan *labelling* pada tiap hasil *cluster*.

Setelah mendapatkan hasil *clustering* untuk masing-masing metode, akan dilakukan evaluasi dengan menghitung *silhouette index*, waktu eksekusi, dan *davies-bouldin index* untuk masing-masing metode.

4. HASIL DAN PEMBAHASAN

4.1 Persiapan Data

Setelah dilakukan pre-processing data di atas, didapatkan sebanyak 8537 tweets yang selanjutnya akan dilakukan penambangan data. Data inilah yang akan diolah pada penelitian ini.

4.2 Term Document Matrix

Dari proses pembobotan TF-IDF dengan Term Document Matrix yang dilakukan, didapatkan bentuk matrix:

	<i>Tweet1</i>	<i>Tweet2</i>	<i>Tweet3</i>	<i>Tweet4</i>	<i>Tweet5</i>	...	<i>Tweet8537</i>
<i>aaa</i>	0.000000	0.0	0.0	0.000000	0.0	...	0.0
<i>abai</i>	0.564896	0.0	0.0	0.000000	0.0	...	0.0
<i>abang</i>	0.000000	0.0	0.0	0.189054	0.0	...	0.0
<i>abdi</i>	0.000000	0.0	0.0	0.000000	0.0	...	0.0
<i>acara</i>	0.000000	0.0	0.0	0.000000	0.0	...	0.0
...

Tabel 1. Term Document Matrix untuk Pembobotan TF-IDF

	<i>Tweet1</i>	<i>Tweet2</i>	<i>Tweet3</i>	<i>Tweet4</i>	<i>Tweet5</i>	...	<i>Tweet8537</i>
<i>aaa</i>	0	0	0	0	0	...	0
<i>abai</i>	1	0	0	0	0	...	0
<i>abang</i>	0	0	0	1	0	...	0
<i>abdi</i>	0	0	0	0	0	...	0
<i>acara</i>	0	0	0	0	0	...	0
...

Tabel 2. Term Document Matrix untuk Pembobotan TF

Hasil pembobotan TF-IDF ini akan digunakan untuk melakukan clustering dengan beberapa metode. Selain itu, akan digunakan dalam melakukan perhitungan nilai silhouette index dan davies-bouldin index.

4.3 K-Means Cluster

Teknik clustering dengan menggunakan metode K-means ini dilakukan dengan memasukkan nilai jumlah cluster sebanyak 10 cluster. Hasil clustering berupa kata yang sering muncul, label, dan jumlah tweet masing-masing cluster dapat dilihat pada tabel di bawah:

<i>No. cluster</i>	<i>Kata paling sering muncul</i>	<i>Label</i>	<i>Jumlah tweet</i>
1	apa, apa apa, vaksin apa, laku, apa vaksin,jadi, kalau, kamu, orang	'Jenis Vaksin'	332
2	Sehat, jangan, patuh, ppkm, tenaga,lupa, tenaga sehat, protokol, protokol sehat, jangan lupa	'Himbauan tetap patuh protokol kessehatan'	543
3	Mau, mau vaksin, kalau, kalau mau, hari, hari mau, vaksin mau, dulu, takut, jadi	'Keinginan vaksin'	396
4	Covid, vaksin covid, vaksinasi covid, door, cepat, herd immunity, immunity, herd, anak, virus	'Herd immunity'	822
5	Jadi, kalau, daftar, sakit, baru, lebih, banyak, bulan, pagi, benar	'Waktu vaksinasi'	4396
6	Sertifikat, sertifikat vaksin, syarat, vaksin sertifikat, kartu, kalau, cek, sertifikat vaksinasi, mau, ktp	'Sertifikat Vaksin'	229
7	Informasi, terima, kasih, terima kasih, kg, informasi vaksin, terima vaksin, fase, iow, pulih nasional	'Informasi vaksin'	271
8	Orang, banyak, orang orang, banyak orang, orang vaksin, semua, jadi, covid, orang vaksinasi, semua orang	'Orang vaksin'	740
9	Dosis, dosis vaksin, dosis pertama, vaksin dosis, pertama, juta, terima, selesai, juta dosis, selesai dosis	'Dosis pertama vaksin'	407
10	Hari, hari vaksin, vaksin hari, hari vaksinasi, vaksinasi hari, pagi, pagi hari, aku, hari hari, moga	'Waktu vaksinasi'	401

Tabel 3. Hasil Clustering dengan Metode K-means

4.4 K-Medoids Cluster

Teknik clustering dengan menggunakan metode K-medoids ini dilakukan dengan memasukkan nilai jumlah cluster sebanyak 10 cluster. Hasil clustering berupa kata yang sering muncul, label, dan jumlah tweet masing-masing cluster dapat dilihat pada tabel di bawah:

No. cluster	Kata paling sering muncul	Label	Jumlah tweet
1	apa mau, nyonya, memberitahu, tuan, kertas, buat apa, digital, email, tindak, rs	'Informasi vaksin'	1
2	Paksa, kata, orang ambil, paksa orang, vaksin paksa, jangan paksa, vaksin takut, dorong, jabat, orang	'Vaksin paksa'	1
3	mau vaksin, mau, zoo, jasa, jatah, jatah vaksin, jatim, jatuh, jatuh sakit, jauh	'Jatah vaksin'	29
4	Dana, vaksinasi laku, akses, tahap, Menteri, beberapa, kota, tahun, bantu, lalu	'Pendanaan vaksin'	1
5	Percaya, warga vaksinasi, paksa, vaksinasi covid, takut, mana, warga, masuk, kalau, covid	'Takut vaksin'	1
6	Akun, sertifikat, beda sama, punya sertifikat, buat, kalau buat, ane, vaksin keluar, vaksin sertifikat, hp	'Sertifikat vaksin'	2
7	Zoo, jejak, jasa, jatah, jatah vaksin, jatim, jatuh, jatuh sakit, jauh, jauh lebih	'Jatah vaksin'	8456
8	Orang, mata, orang orang, banyak orang, orang ragu, perintah vaksin, tinggal orang, selidik, banyak, vaksin mati	'Ragu vaksin'	1
9	vaksin pertama, dosis vaksin, pertama, dosis, orang, jawa timur, jawa tengah, jawa barat, jawa, jauh lebih	'Dosis pertama vaksin'	5
10	hari vaksin, hari, jaya, jauh, jasa, jatah, jatah vaksin, jatim, jatuh, jatuh sakit	'Jatah vaksin'	40

Tabel 4. Hasil Clustering dengan Metode K-medoids

4.5 GSDMM Cluster

Teknik clustering dengan menggunakan metode GSDMM ini dilakukan dengan memasukkan nilai jumlah cluster sebanyak 10 cluster. Hasil clustering berupa kata yang sering muncul, label, dan jumlah tweet masing-masing cluster dapat dilihat pada tabel di bawah:

No. cluster	Kata paling sering muncul	Label	Jumlah tweet
1	Hari, mau , pagi, sakit, apa, jadi, bulan, dosis, pertama, kalau	'Perintah vaksin dosis pertama'	662
2	Sehat, covid, perintah, orang, masyarakat, jadi, pandemic, ppkm, negara, banyak	'Alasan vaksin'	1549
3	Orang, covid, mau, apa, jadi, banyak, kalau, sakit, jangan, benar	'Alasan vaksin'	237
4	Covid, orang, dosis, lebih, kasus, jadi, terima, hari, bulan, banyak	'kasus covid'	1828
5	Mau, sertifikat, daftar, ktp, kalau, kartu, jadi, hari, orang, mana	'sertifikat vaksin'	1196
6	Covid, anak, masyarakat, hari , kota, program, laksana, warga, cepat, laku	'pelaksanaan vaksin'	1381
7	Ni, mau, orang, kalau, nak, covid, rm, dah, buat, tu	'lain-lain'	153
8	Astrazeneca, orang, Menteri, sarah_gilbert, anak, temu, siapa, bilang, baru, dunia	'vaksin astrazeneca'	165
9	Dosis, hari, vaccine, ppv, beri covid, mulai, juli, Indonesia, people	'pemberian dosis vaksin'	125
10	Informasi, covid, fase, lokasi, kabupaten, taman, daftar, dun , juli, kantor	'informasi vaksin'	1241

Tabel 5. Hasil Clustering dengan Metode GSDMM

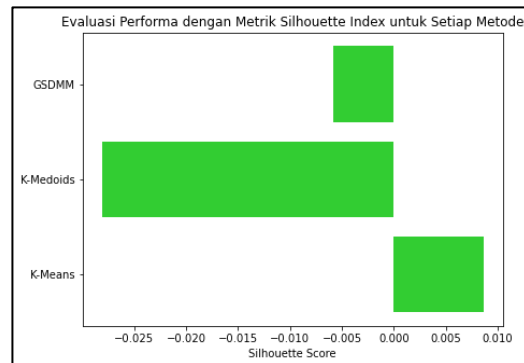
4.6 Analisis Hasil Clustering

1. Silhouette index

Salah satu perhitungan untuk menentukan kualitas suatu hasil dari clustering adalah perhitungan nilai silhouette index. Jika nilai Silhouette Index semakin mendekati 1, maka cluster tersebut dinyatakan semakin baik. Hasil perhitungan silhouette index untuk masing-masing metode adalah sebagai berikut:

Metode Clustering	Nilai Silhouette Index
K-Means	0.008611686544428203
K-Medoids	-0.028096307637543467
GSDMM	-0.005841526355796398

Tabel 6. Nilai Silhouette Index



Gambar 1. Grafik Performa Metode dengan Silhouette Index

Dari hasil perhitungan nilai silhouette index didapatkan k-means sebagai clustering terbaik karena nilainya yang paling mendekati 1 yaitu 0.00861 dibandingkan dengan metode k-medoids dan GSDMM.

2. Waktu Eksekusi

Perhitungan waktu eksekusi dilakukan dengan 10 cluster untuk semua metode. Hasil dari perhitungan waktu eksekusi clustering adalah:

Metode Clustering	Nilai Silhouette Index
K-Means	14.919600248336792
K-Medoids	6.917362928390503
GSDMM	191.8668143749237

Tabel 7. Waktu Esekusi Masing-Masing Metode

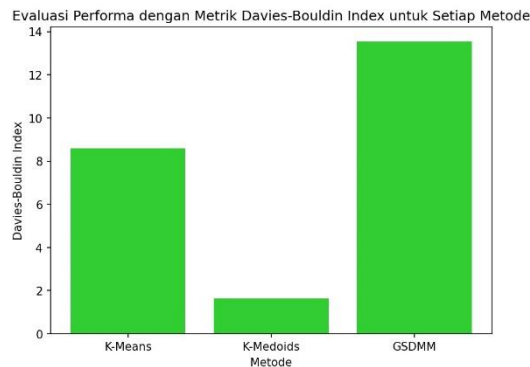
Dengan melihat hasil perhitungan waktu eksekusi didapatkan metode k-medoids dengan waktu eksekusi tercepat yaitu 6.92 sekon, kemudian ada k-means dengan waktu eksekusi 14.92 sekon, dan terakhir metode GSDMM dengan waktu eksekusi 191.87 sekon.

3. Davies-Bouldin Index

Selain perhitungan Silhouette Index untuk menguji tingkat validitas suatu cluster, ada metode lain yaitu perhitungan Davies-Bouldin Index. Metode ini mengevaluasi cluster dengan melihat kuantitas dan kedekatan antar data hasil cluster. Semakin kecil nilai Davies-Bouldin Index, maka semakin baik cluster yang diperoleh dari metode yang digunakan untuk Teknik clustering. Pada penelitian ini, hasil dari perhitungan nilai Davies-Bouldin Index dapat dilihat pada tabel berikut:

Metode Clustering	Nilai Davies-Bouldin Index
K-Means	8.592422274737585
K-Medoids	1.6277003785404003
GSDMM	13.547077449925634

Tabel 8. Nilai Davies-Bouldian Index



Gambar 2. Grafik Evaluasi Performa dengan DBI

Dari hasil perhitungan nilai DBI di atas, didapatkan hasil metode K-medoids dengan performa yang baik dikarenakan nilai DBI nya yang paling rendah dengan nilai DBI 1.6277. Metode dengan nilai DBI tertinggi yaitu metode GSDMM menjadi metode dengan performa paling rendah. Sedangkan metode dengan nilai DBI sedang yaitu metode K-means.

5. KESIMPULAN

Dari hasil penelitian penambahan data dengan Teknik clustering pada data Tweet Vaksinasi Indonesia dengan metode K-Means, K-Medoids, dan GSDMM yang telah dilakukan, didapatkan kecenderungan pembahasan masyarakat mengenai vaksin adalah topik 'jenis vaksin'. Selain itu, dari hasil evaluasi dengan menggunakan tiga parameter yaitu, Silhoutte Index, Davies-Bouldin Index, dan waktu eksekusi didapatkan metode K-means dengan hasil terbaik karena memiliki nilai stabil pada tiga parameter tersebut (Silhoutte index posisi pertama, Davies-Bouldin Index posisi kedua, dan waktu eksekusi pada posisi kedua).

Pada penelitian berikutnya diharapkan dapat menambah metode baru dalam Teknik clustering untuk dibandingkan dengan metode lain agar hasil dari penelitian dapat dijadikan referensi baru. Selain itu, masih banyak metode yang dapat digunakan untuk mengevaluasi hasil clustering. Beberapa metode yang dapat digunakan adalah perhitungan Dunn Index, Calinski-Harabasz Index, Partition Coefficient Index, dan masih banyak metode lainnya.