



PENGELOMPOKAN RUMAH SAKIT DI JAKARTA MENGGUNAKAN MODEL DBSCAN, GAUSSIAN MIXTURE, DAN HIERARCHICAL CLUSTERING

Karli Eka Setiawan¹, Afdhal Kurniawan²

^{1,2}Computer Science Department, School of Computer Science, Bina Nusantara University
Jakarta Barat, DKI Jakarta, Indonesia 11480

karli.setiawan@binus.ac.id, afdhal.kurniawan@binus.ac.id

Abstract

It should be recognized that after the COVID-19 pandemic, the distribution of hospital facilities should be a primary concern in meeting society's fundamental rights to local and national governments. Therefore, it is necessary to analyze data that describes the current conditions regarding the distribution of hospital facilities, especially in Jakarta. Data analysis using a machine learning technique can provide several benefits, such as understanding the distribution of health workers, providing insight into human resource planning, monitoring hospital performance, and even becoming a reference for future targets by predicting the need for health workers. The research proposed three models of unsupervised learning, such as the density-based spatial clustering of applications with noise algorithm (DBSCAN), the Gaussian mixture model, and the agglomerative hierarchical model, to cluster a database of hospitals in Jakarta, Indonesia, that contains information on the number of medical workers in the hospital and its bed facilities. This data set was obtained from a web scraping process on the Ministry of Health of the Republic of Indonesia website in 2022. The comparison of the third unsupervised model showed that the Gaussian Mixture model produced the smallest Davies-Bouldin value with a value of 0.645713. This study is an extension of our previous investigation, in which the study is, as far as we know, the first to discuss the classification of the data on the list of hospitals in Jakarta based on information on the number of medical personnel and the number of bed facilities that make up the contribution of this research.

Keywords: Agglomerative Hierarchical, DBSCAN, Gaussian Mixture, Hospital Clustering, unsupervised learning

Abstrak

Perlu disadari bahwa setelah mengalami pandemi COVID-19, distribusi persebaran fasilitas rumah sakit harus menjadi perhatian utama dalam memenuhi hak-hak dasar dari masyarakat oleh pemerintah tingkat daerah dan tingkat nasional. Oleh karena itu diperlukan sebuah analisis data yang menggambarkan kondisi terkini mengenai persebaran fasilitas rumah sakit khususnya di Jakarta. Analisis Data dengan pendekatan teknik machine learning mampu memberikan beberapa manfaat seperti mendapatkan manfaat dalam memberikan pemahaman mengenai distribusi tenaga kesehatan, memberikan wawasan dalam rangka perencanaan sumber daya manusia, melakukan pemantauan kinerja rumah sakit, dan bahkan menjadi acuan target di masa depan dengan prediksi kebutuhan tenaga kesehatan. Penelitian ini mengajukan tiga buah model *unsupervised learning*, seperti model *Density-based spatial clustering of applications with noise algorithm* (DBSCAN), model *Gaussian Mixture*, dan model *Agglomerative Hierarchical* untuk melakukan pengelompokan sekumpulan data rumah sakit se-Jakarta, Indonesia yang berisi informasi data jumlah berbagai tenaga kerja medis di rumah sakit tersebut dan fasilitas tempat tidurnya. *Dataset* ini diperoleh dari proses *web scraping* pada *website* Kementerian Kesehatan Republik Indonesia pada tahun 2022. Dari perbandingan ketiga model *unsupervised*, diperoleh bahwa model *Gaussian Mixture* menghasilkan nilai *Davies Bouldin* terkecil dengan nilai 0,645713. Penelitian ini merupakan lanjutan dari penelitian kami sebelumnya, dimana penelitian tersebut menurut sepengetahuan kami merupakan penelitian pertama yang membahas pengelompokan data daftar rumah sakit se-Jakarta berdasarkan informasi data jumlah tenaga medis dan jumlah fasilitas tempat tidur yang menjadikan kontribusi dari penelitian ini.

Kata kunci: Agglomerative Hierarchical, DBSCAN, Gaussian Mixture, pengelompokan rumah sakit, *unsupervised learning*

1. PENDAHULUAN

Pengembangan di sektor kesehatan kini menjadi bagian dari perhatian pemerintah pusat dan daerah dengan tindakan

nyata berupa menyediakan fasilitas kesehatan yang memadai dan meningkatkan angka harapan hidup Masyarakat Indonesia karena Kesehatan Masyarakat

merupakan salah satu dari hak-hak dasar dari masyarakat yang harus direalisasikan [1]. Pengumpulan data dan analisis performa dari fasilitas Kesehatan dalam suatu negara merupakan hal yang penting untuk kebijakan pemerintah di masa depan, khususnya kebijakan terkait Kesehatan Masyarakat. Indonesia negara yang memiliki jumlah penduduk yang sangat banyak yaitu sekitar lebih dari 200 juta penduduk yang tentunya memiliki permasalahan dalam melakukan distribusi persebaran fasilitas Kesehatan dan tenaga ahli Kesehatan seperti dokter, perawat, bidan dan lainnya yang dimana saat ini mayoritas tersebar di daerah kota-kota besar [2]. Hal tersebut dapat terlihat ketika Indonesia menghadapi pandemi COVID-19 dimana tenaga medis yang tersedia tidak mencukupi dan meningkatnya permintaan penggunaan fasilitas Kesehatan oleh para pasien [3]. Saat ini akses untuk memperoleh data fasilitas rumah sakit dengan informasi detail di Indonesia sangat terbatas. Oleh karena itu menjadi pekerjaan yang sangat berat bagi Masyarakat untuk menentukan level dari rumah sakit yang ada di Indonesia. Terlebih lagi data terkait berapa jumlah dokter dan dokter spesialis, perawat, bidan, jumlah tempat tidur, fasilitas *intensive care unit* (ICU), dan layanan unit gawat darurat di sebuah rumah sakit tertentu tersebar di *website* masing-masing setiap rumah sakit. Oleh karena itu penelitian ini melanjutkan penelitian kami sebelumnya yang setelah memiliki *dataset* rumah sakit se-Jakarta pada tahun 2022 dengan menerapkan model *unsupervised machine learning* untuk melakukan pengelompokan data rumah sakit se-Jakarta. Diharapkan dengan adanya riset ini maka dapat membantu pemerintah pusat maupun pemerintah daerah dalam membuat keputusan terkait kebijakan dibidang kesehatan dan distribusi tenaga kesehatan.

Kontribusi dari penelitian ini adalah komparasi tiga buah model *unsupervised learning* yaitu DBSCAN, GMM, dan *Hierarchical* untuk melakukan pengelompokan data rumah sakit se-Jakarta. Sepengetahuan kami sebelumnya belum ada pembahasan riset *paper* yang membahas pengelompokan rumah sakit di Indonesia berdasarkan tenaga kerja medis dan jumlah fasilitas tempat tidur kecuali penelitian kami sebelumnya [4].

Dalam bidang medis, pendekatan *unsupervised learning* untuk keperluan pengelompokan data terbukti sangat mendatangkan banyak manfaat dengan mengungkapkan fakta-fakta yang tidak dapat dilihat oleh mata manusia jika tidak dari kumpulan data yang sangat besar jumlahnya [5]. Pada penelitian sebelumnya pendekatan dengan *unsupervised learning* telah digunakan untuk keperluan mendiagnosis berbagai penyakit seperti kanker payudara, Parkinson, sakit kepala, penyakit mental, penyakit jantung, diabetes, dan banyak lainnya. Seorang peneliti mempelajari pendeteksian penyakit Alzheimer dengan pendekatan pengelompokan *unsupervised* dengan menggunakan model *K-Means*. *K-Means-Mode*, *clustering multilayer* dan *hierarchical agglomerative* [5]. Terdapat penelitian lainnya yang melakukan kalsterisasi dibidang medis dengan

pengelompokan data dari banyak kelompok penderita *tuberculosis* berdasarkan banyak atribut seperti usia, jenis kelamin, status infeksi *human immunodeficiency virus*, riwayat penyakit *diabetes mellitus*, hasil tes sinar X, dan hasil *Rapid test* dengan menggunakan model *K-Means* dan *Fuzzy C-Means* [6].

Terdapat sebuah penelitian tentang teknik *clustering* yang dilakukan oleh Ghosal dan beberapa peneliti lainnya, mereka menyebutkan bahwa model pengelompokan *Hierarchical* sudah digunakan di berbagai bidang seperti pendidikan, penggambaran pola di sosial media, mesin pencari *web*, penanganan informasi sampingan dan *metadata* di dalam sebuah dokumen, pengurangan konsumsi energi dalam sensor lintas jaringan sensor, pengelompokan perilaku *malware*, dan segmentasi gambar dengan model *hierarchical* dan model pengelompokan DBSCAN sudah digunakan pada sistem regulator anti pencucian uang, pemaksimalan keuntungan dan ekspansi pasar, manajemen keinginan pelanggan, manajemen lalu lintas jaringan, penyiapan fasilitas seperti bank, sekolah, rumah sakit dan sebagainya, dan instalasi ATM dan *e-corners* pada tempat strategis [7].

2. METODE PENELITIAN

2.1 Dataset

Dataset yang dieksplorasi pada studi ini merupakan *dataset* yang kami miliki dari penelitian kami sebelumnya pada tahun 2022 [4]. *Dataset* tersebut diperoleh dari proses web scraping dengan menggunakan teknik *parsing* HTML dari *website* yang bernama “Badan PPSDM Kesehatan Informasi SDM Kesehatan Kementerian Kesehatan Republik Indonesia versi 4.0” yang merupakan dibawah pertanggungjawaban Kementerian Kesehatan Republik Indonesia. Jika dicek kembali *website* tersebut pada tahun 2023, tampilan *website* tersebut sudah banyak berubah dan tidak dapat diperoleh kembali *dataset* yang sama persis penelitian ini miliki. Proses Web Scraping yang penelitian sebelumnya kami lakukan merupakan sebuah cara yang legal dan sah karena data tersebut bersifat publik. Proses *Web Scraping* merupakan proses yang cukup efektif dan legal untuk mengumpulkan dan melakukan ekstraksi data secara otomatis yang bekerja seperti *bot scraper* pada jaringan internet dari *website*, kemudian menyusunnya menjadi basis data terstruktur berdasarkan hasil proses dari *web scraping* [8].

Dataset yang diteliti pada eksperimen ini merupakan kumpulan dari berbagai macam tabel seperti tabel dokter spesialis, psikologi klinis, perawat, bidan, farmasi, kesehatan masyarakat, kesehatan lingkungan, nutrisioonis, terapi fisik, teknisi medis, dan biomedis rekayasa. Secara keseluruhan *dataset* ini berisi data rumah sakit se-DKI Jakarta dengan detail jumlah tenaga kerja medis dan jumlah kapasitas tempat tidurnya dengan detail *dataset* yang dijelaskan secara rinci pada Tabel 1.

Tabel 1. Informasi yang Terkandung dalam Dataset

Nama Tabel	Deskripsi dari Jenis-jenis Tenaga Medis yang tersedia dalam Tabel	Jumlah tipe unik
Dokter Spesialis	Dokter umum, Dokter gigi, 42 jenis dari dokter spesialis yang berbeda, 10 jenis dokter gigi spesialis yang berbeda, dan 1 dokter sub spesialis	55
Psikologi Klinis	Psikolog Klinis	1
Perawat	Ners, Perawat (Non Ners), Perawat Kesehatan Anak, Perawat Maternitas, Perawat Medis Bedah, Perawat Geriatri – Lansia, Perawat Kesehatan Jiwa, Perawat Komunitas, Perawat Lainnya, dan Perawat (SPK)	10
Bidan	Bidan Klinik, Bidan Desa, Bidan Pendidik, Bidan Umum, dan Bidan (P2B, Bidan A & Bidan C)	5
Farmasi	Apoteker, Ahli Farmasi bukan Apoteker, Asisten Apoteker, Farmasi Analis, dan Farmasi (SMF)	5
Kesehatan Masyarakat	Ahli Epidemiologi Kesehatan, Promosi Kesehatan, Pakar Perilaku, Pakar Keselamatan dan Kesehatan Kerja, Pengelola dan Kebijakan Kesehatan, Statistik Kesehatan dan Demografi, Ahli Kesehatan Reproduksi dan Peranan Keluarga, Ahli Teknologi Informasi Kesehatan, Kesehatan Masyarakat Umum, Ahli Sanitasi dan Kebersihan Lingkungan, Pakar Entomolog Kesehatan, Peneliti dan ahli Mikrobiologi Medis, dan Kesehatan Lingkungan (SPPH)	9
Kesehatan Lingkungan	Sanitasi Lingkungan, Entomolog Kesehatan, Mikrobiolog Kesehatan, dan Kesehatan Lingkungan (SPPH)	4
Nutrisionis	Nutrisionis, Dietisien, dan Gizi (SPAG)	3
Terapi Fisik	Ahli Fisioterapi, Terapis Okupasi, Ahli Terapis Wicara, dan Pengobatan Akupunktur	4
Teknisi Medis	Tenaga Medis Rekam dan Informasi Kesehatan, Spesialis Teknisi Kardiovaskular, Spesialis Teknisi Pelayanan Darah, Spesialis Penglihatan, Teknisi Dental, Penata Anestesiologi, Terapis Kedokteran Gigi, Ahli Audiologi, dan Terapis Gigi dan Mulut/Perawat Gigi (SPRG)	9
Biomedis Rekayasa	Ahli Radiografi, Ahli Elektromedis, Tenaga Laboratorium Kesehatan (Analis Kesehatan), Spesialis Fisika Medis, Ahli Radioterapi, Ahli Ortotik Prostetik, dan Praktisi Analis Kesehatan (SMAK)	7

2.2 Principal Component Analysis (PCA)

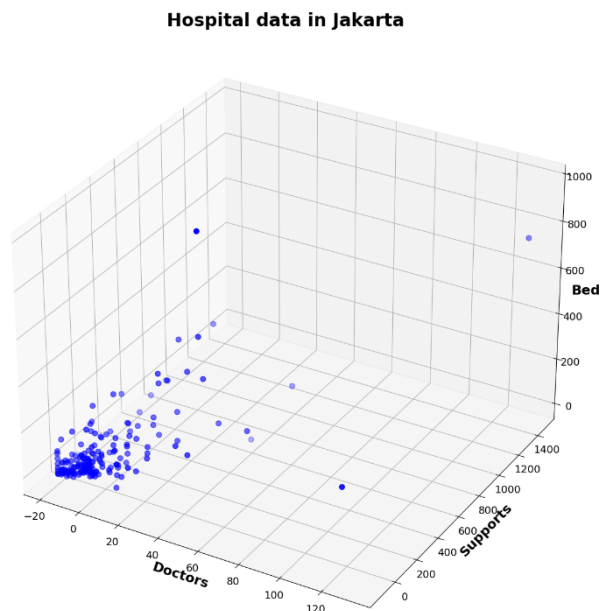
Merupakan sebuah pekerjaan yang sangat menantang untuk mengilustrasikan dimensi yang sangat besar yang terlebih lagi terdiri ratusan dimensi. *Principal Component Analysis* (PCA) yang merupakan metode matematika digunakan dalam penelitian ini untuk menurunkan jumlah dimensi sehingga data yang akan dikelompokkan dapat divisualisasikan dengan mudah untuk keperluan analisis yaitu menjadi tiga dimensi untuk penelitian ini [9].

Penelitian ini mentransformasikan 55 dimensi dari berbagai jenis dokter menjadi satu buah dimensi dokter atau sumbu Dokter, 57 dimensi dari berbagai jenis tenaga medis bukan dokter seperti psikologi klinis, keperawatan, kebidanan, farmasi, kesehatan masyarakat, kesehatan lingkungan, gizi, terapi fisik, teknisi medis, dan teknik biomedis menjadi satu buah dimensi pendukung atau sumbu pendukung.

$$x_n = W^t(t_n - \bar{t}) \quad (1)$$

Hasil dari penurunan dimensi diilustrasikan dalam persamaan (1) dengan dinotasikan dengan simbol x_n [10]. Simbol dari W merepresentasikan sumbu utama dengan $W = (w_1, w_2, \dots, w_q)$ dimana q merupakan sumbu ortonormal yang diperoleh ketika varians di bawah proyeksi maksimal, t_n mewakili sebuah set dari *vector* data d -

dimensi yang diobservasi dengan $n \in \{1, \dots, N\}$, dan \bar{t} merepresentasikan nilai rata-rata dari sampel.



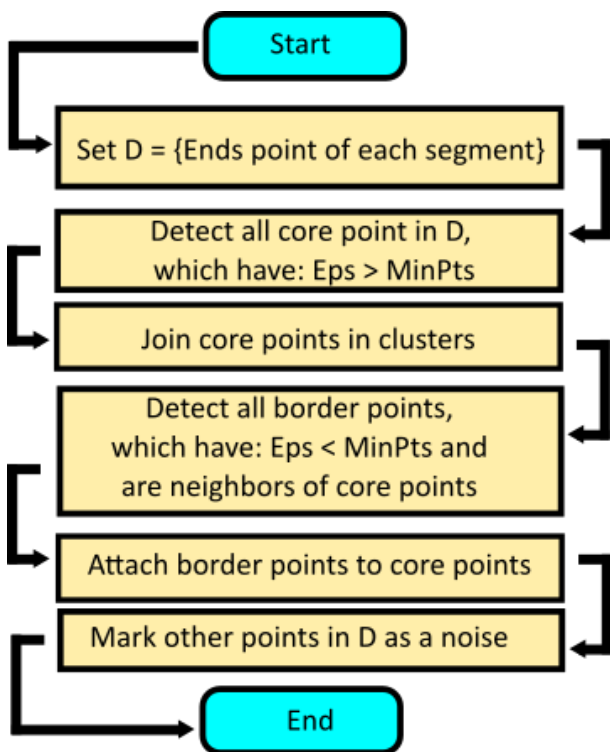
Gambar 1. Visualisasi Dataset Rumah Sakit di Jakarta dalam 3 Dimensi

Hasil dari proses PCA diperoleh nilai dimensi Dokter akan diperoleh 132,13 untuk nilai maksimum, -17,53 untuk nilai

minimum, dan $5,95 \times 10^{-16}$ untuk nilai rata-rata. Sementara untuk dimensi Pendukung diperoleh nilai 1.456,77 untuk nilai maksimum, -111,82 untuk nilai minimum, dan $4,6 \times 10^{-15}$ untuk nilai rata-rata. Sementara untuk dimensi fasilitas tempat tidur tidak melalui proses PCA sehingga datanya berupa data aktual dengan 968 sebagai nilai maksimum, 0 sebagai nilai minimum, dan 97,14 sebagai nilai rata-rata. Sehingga *dataset* dari yang digunakan dalam penelitian ini dapat divisualisasikan pada Gambar 1.

2.3 Model DBSCAN

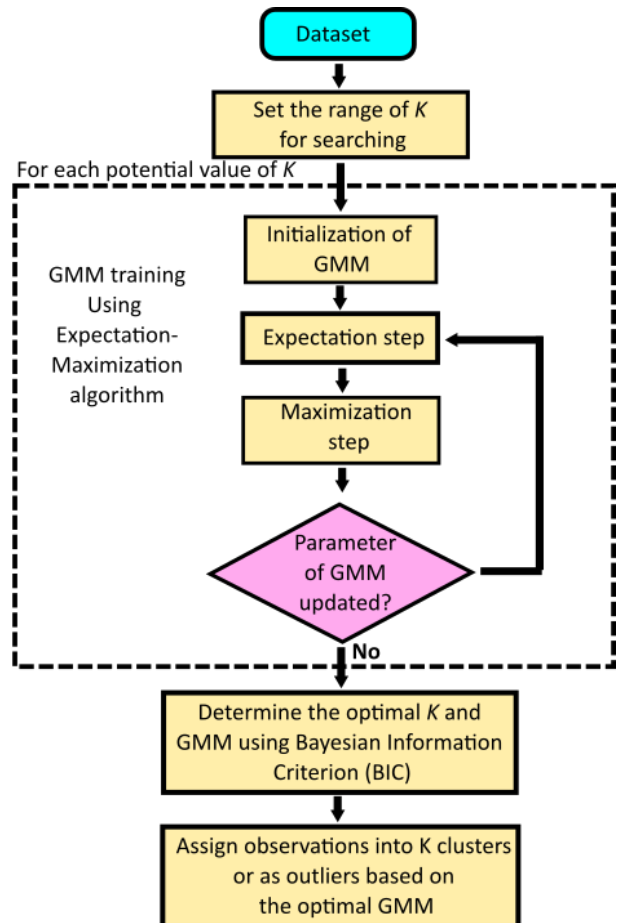
Density-based spatial clustering of applications with noise algorithm (DBSCAN) merupakan bagian dari salah satu dari pendekatan *unsupervised learning* yang merupakan algoritma *clustering* berbasis kepadatan [11], [12]. DBSCAN dapat membentuk *cluster* yang melalui penentuan area berdasarkan kepadatan tinggi yang dipisahkan dengan kepadatan yang rendah berdasarkan kepadatan cluster. DBSCAN dapat bekerja dengan baik untuk *dataset* yang berbentuk apa pun dalam *database* spasial dan efektif untuk mengelompokkan data yang sangat sulit sekalipun. Alur kerja DBSCAN dapat dilihat pada Gambar 2 dimana DBSCAN dimulai dengan titik acak. Notasikan *D* sebagai *dataset*, *Eps* sebagai radius dari lingkungan yang berdekatan dari titik data yang dipertimbangkan, dan *MinPts* sebagai jumlah minimum titik data yang berdekatan di wilayah tertentu.



Gambar 2. Flowchart dari DBSCAN

2.4 Model Gaussian Mixture (GMM)

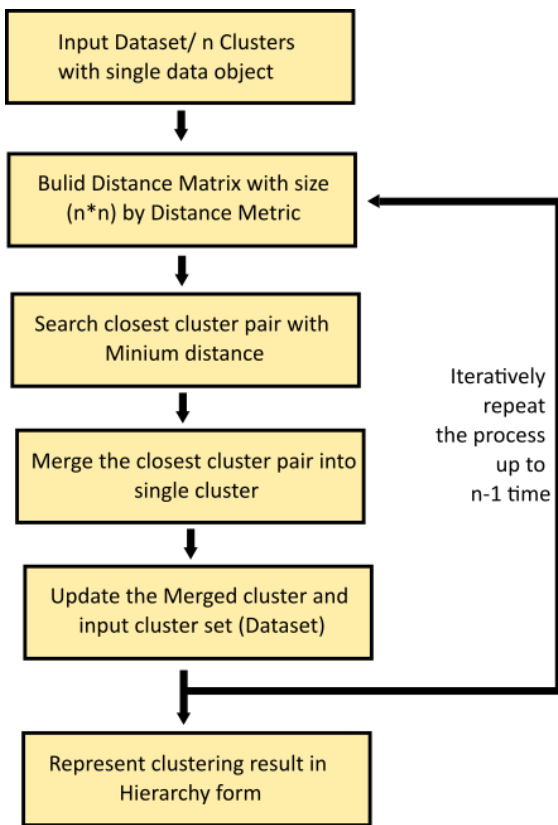
Gaussian Mixture Model (GMM) merupakan model *clustering* berbasis probabilitastas yang memanfaatkan nilai bobot dari kombinasi beberapa distribusi normal sehingga model ini sering disebut sebagai komponen campuran (*mixture component*) [13]. *Flowchart* dari GMM dapat dilihat pada Gambar 3. Algoritma *clustering* ini dapat mengidentifikasi dan menghilangkan data observasi yang memiliki probabilitas rendah dalam komponen campuran yang didasarkan pada *outliers* untuk meningkatkan kekokohan dalam hasil *clustering*.



Gambar 3. Flowchart dari GMM

2.5 Model Agglomerative Hierarchical

Model *agglomerative hierarchical cluster* (AHC) merupakan metode *clustering* hierarki *bottom-up* yang menggabungkan *N cluster* menjadi sebuah *cluster* tunggal [14]. Langkah-langkah dasar dari model AHC dapat dilihat pada Gambar 4 [15].



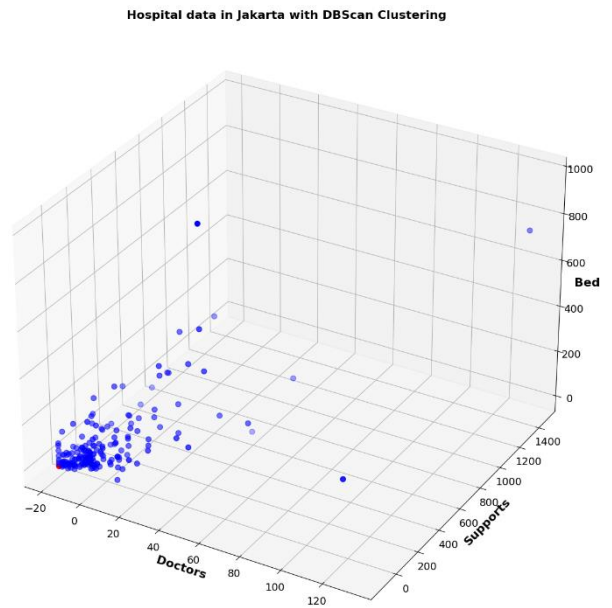
Gambar 4. Alur dari Agglomerative Hierarchical

3. HASIL DAN PEMBAHASAN

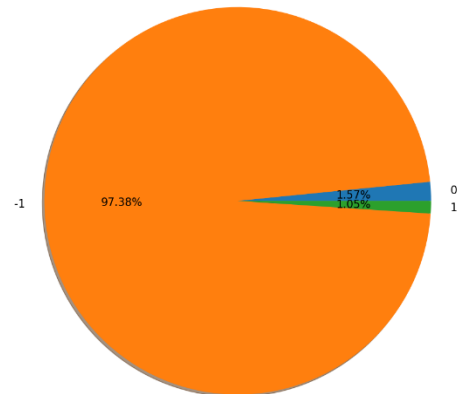
Eksperimen *clustering* dengan data rumah sakit di wilayah Provinsi Jakarta dilakukan dengan menggunakan model *clustering* DBSCAN, GMM, dan AHC yang tersedia pada *library sklearn*.

3.1 Pengelompokan dengan Model DBSCAN

Untuk mencari nilai parameter *eps* yang terbaik penelitian ini mencoba berbagai kemungkinan dalam rentang $0,3 \leq eps \leq 0.9$ dengan diperoleh nilai terbaiknya adalah *eps* = 0,3 dan diperoleh nilai *silhouette* sebesar 0,473. Dari hasil yang eksperimen dengan menggunakan model pengelompokan DBSCAN diperoleh bahwa dari total 191 rumah sakit di Jakarta dapat dikelompokkan menjadi 3 kelompok saja dimana 97,38% dikelompokkan menjadi *cluster* -1, sementara hanya 1,57% dikelompokkan menjadi *cluster* 0 dan 1,05% dikelompokkan menjadi *cluster* 1 seperti pada Gambar 6. Hasil pengelompokan dengan model DBSCAN dapat dilihat pada Gambar 5.



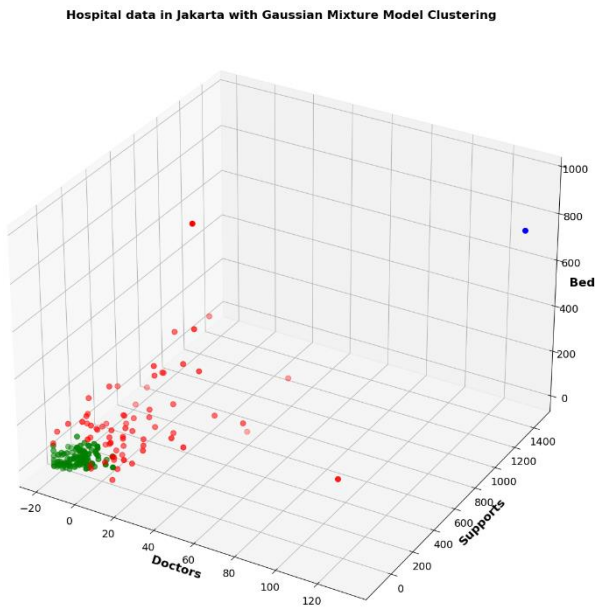
Gambar 5. Hasil pengelompokan Rumah Sakit di Jakarta dengan Model DBSCAN



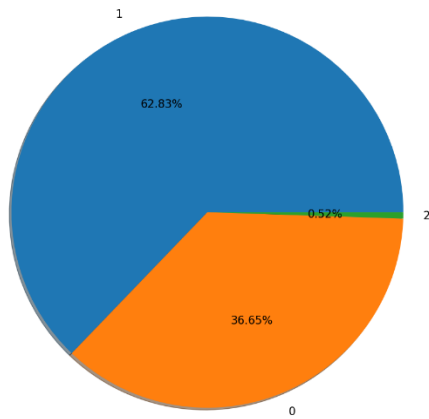
Gambar 6. Proporsi dari Hasil Pengelompokan Rumah Sakit di Jakarta dengan Model DBSCAN

3.2 Pengelompokan dengan Model Gaussian Mixture

Pada penelitian ini untuk memperoleh jumlah *cluster* yang optimal untuk model GMM ini, maka penelitian ini melakukan pencarian dengan menggunakan *elbow method* mencoba semua kemungkinan jumlah *cluster* dalam rentang $2 \leq \text{jumlah klaster} \leq 10$. Diperoleh jumlah kluster yang optimal sebanyak tiga buah *cluster* dengan nilai *silhouette* 0,441. Ilustrasi dari persebaran data rumah sakit di Jakarta dengan tiga buah *cluster* dapat dilihat pada Gambar 7. Distribusi persebaran data rumah sakit se-Jakarta dengan jumlah tiga *cluster* dapat dilihat pada Gambar 8, dimana dari total 191 rumah sakit tersebar sebanyak 62,83% masuk dalam *cluster* 1, sebanyak 36,65% masuk dalam *cluster* 0, dan sebanyak 0.52% masuk dalam *cluster* 2.



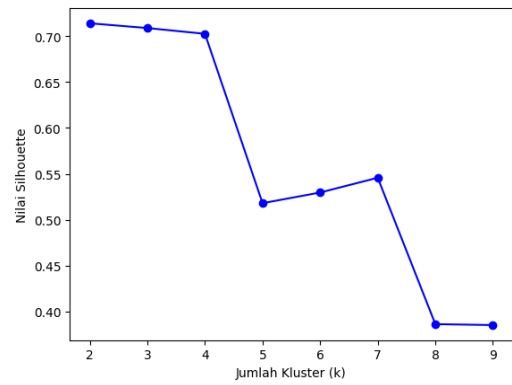
Gambar 7. Hasil Pengelompokan Rumah Sakit di Jakarta dengan Model Gaussian Mixture



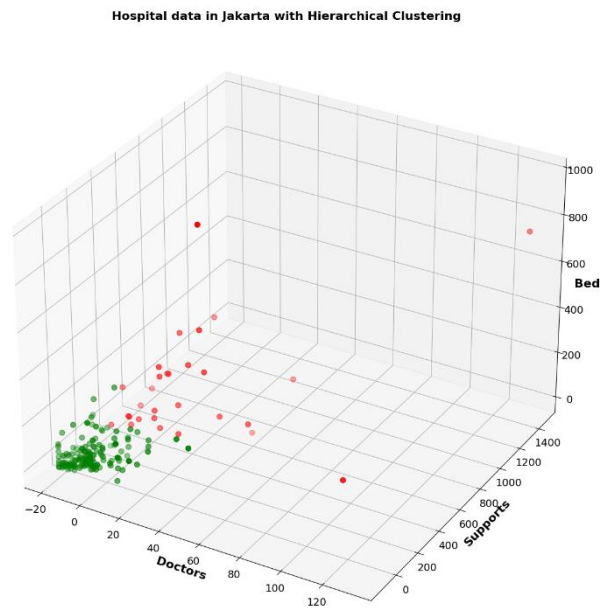
Gambar 8. Proporsi dari Hasil Pengelompokan Rumah Sakit di Jakarta dengan Model Gaussian Mixture

3.3 Pengelompokan dengan Model Hierarchical

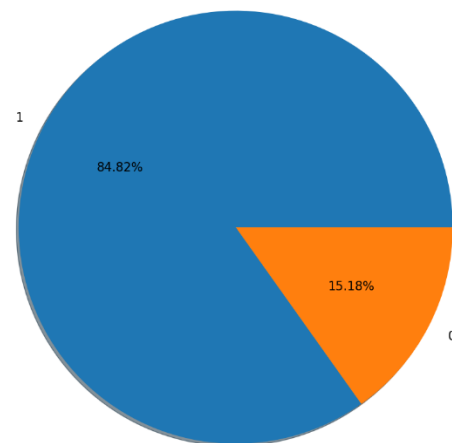
Pada eksperimen dengan model *Agglomerative Hierarchical* ini, peneliti melakukan pencarian jumlah *cluster* yang optimal dengan menggunakan *elbow method* dengan pencarian dari dua buah *cluster* hingga sembilan buah *cluster*, sehingga diperoleh hasil seperti di Gambar 9. Namun peneliti menggunakan bantuan pencarian sudut *elbow* dengan *library KneeLocator*, sehingga diperoleh jumlah *cluster* yang tepat adalah dua buah *cluster*. Visualisasi dari persebaran data rumah sakit di Jakarta dengan dua buah *cluster* dapat dilihat pada Gambar 10. Dari total 191 jumlah rumah sakit di Jakarta, sebanyak 84,82% masuk dalam kategori *cluster* 1 dan 15,18% masuk ke dalam kategori *cluster* 0 yang divisualisasikan di Gambar 11.



Gambar 9. Nilai Silhouette untuk Pencarian Jumlah Cluster yang Optimal



Gambar 10. Hasil Pengelompokan Rumah Sakit di Jakarta dengan Model Hierarchical



Gambar 11. Proporsi dari Hasil Pengelompokan Rumah Sakit di Jakarta dengan Model Hierarchical

3.4 Diskusi

Untuk membandingkan ketiga model *unsupervised* dalam melakukan pengelompokan data rumah sakit di Jakarta, penelitian ini menggunakan *index Davies Bouldin*. Index dengan *Davies Bouldin* digunakan untuk mengukur seberapa baik proses pengelompokan telah dilakukan dengan semakin rendah nilai *index* tersebut maka semakin baik hasil pengelompokan. Dengan menghitung rasio jarak antara intra-cluster dan jarak inter-cluster diperoleh bahwa model terbaik dalam kasus penelitian ini adalah model *clustering* dengan *Gaussian Mixture* seperti pada Tabel 2.

Tabel 2. Perbandingan Ketiga Model *Clustering*

<i>Clustering Model</i>	<i>Davies Bouldin Score</i>
<i>DBScan Clustering</i>	1,2258397217676922
<i>Gaussian Mixture Clustering</i>	0,6457134069852491
<i>Hierarchical Clustering</i>	0,6680066216696465

4. KESIMPULAN

Penerapan 3 model *clustering* yaitu dengan DBSCAN, *Gaussian Mixture*, dan *Agglomerative Hierarchical* dapat menghasilkan bentuk berbagai *cluster* yang berbeda-beda baik itu dari segi luas *cluster* maupun jumlah *cluster*. Dengan menggunakan DBSCAN diperoleh tiga buah *cluster* dengan proporsi 97,38%, 1,57%, dan 1,05%. Penggunaan model *Gaussian Mixture* diperoleh tiga buah *cluster* juga dengan proporsi 62,83%, 36,65%, dan 0.52%. Sementara jika dengan menggunakan model *clustering Agglomerative Hierarchical* dihasilkan dua buah *cluster* saja dengan proporsi 84,82% dan 15,18%. Berdasarkan perbandingan nilai *Davies Bouldin*, didapatkan bahwa model GMM menghasilkan nilai terendah dengan perbandingan DBSCAN menghasilkan nilai *Davies Bouldin* 89.84% lebih besar dibandingkan dengan GMM dan *agglomerative Hierarchical* menghasilkan nilai *Davies Bouldin* 3.45% lebih besar dibandingkan dengan GMM, sehingga hal ini menunjukkan bahwa model GMM adalah yang terbaik. Bagaimanapun dalam penelitian *unsupervised learning* sangatlah bergantung pada jenis *dataset* dalam kesamaan data, sehingga tidak ada jawaban yang tepat salah maupun tepat benar.

Untuk penelitian selanjutnya dimasa depan, peneliti merencanakan untuk melakukan pengumpulan *dataset* rumah sakit yang lebih luas lagi yaitu mencakup satu Negara Kesatuan Republik Indonesia baik melalui cara *web scraping* dari data publik ataupun mengajukan permintaan ke Kementerian Kesehatan Republik Indonesia.

DAFTAR PUSTAKA

[1] H. Bakri, "The planning of community health center in Indonesia," *European Journal of Research and Reflection in Management Sciences*, vol. 6, no. 3, pp. 12–18, 2018, [Online]. Available:

<http://www.idpublications.org/wp-content/uploads/2018/09/Full-Paper-THE-PLANNING-OF-COMMUNITY-HEALTH-CENTER-IN-INDONESIA.pdf>

- [2] M. H. Basri, D. A. Farmaciawaty, I. N. Rachmania, F. B. Widjaja, and A. F. Wiyana, "Healthcare Ecosystem Mapping in Primary Care: A Case Study in West Java, Indonesia," *Proceedings of the International Conference on Business and Engineering Management (ICONBEM 2021)*, vol. 177, pp. 184–190, 2021, doi: 10.2991/aebmr.k.210522.025.
- [3] Y. Mahendradhata *et al.*, "The Capacity of the Indonesian Healthcare System to Respond to COVID-19," *Front Public Health*, vol. 9, no. July, pp. 1–9, 2021, doi: 10.3389/fpubh.2021.649819.
- [4] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, "Clustering models for hospitals in Jakarta using fuzzy c-means and k-means," *Procedia Comput Sci*, vol. 216, pp. 356–363, 2023, doi: 10.1016/j.procs.2022.12.146.
- [5] H. Alashwal, M. El Halaby, J. J. Crouse, A. Abdalla, and A. A. Moustafa, "The application of unsupervised clustering methods to Alzheimer's disease," *Front Comput Neurosci*, vol. 13, no. May, pp. 1–9, 2019, doi: 10.3389/fncom.2019.00031.
- [6] E. M. Rochman, Miswanto, and H. Suprajitno, "COMPARISON OF CLUSTERING IN TUBERCULOSIS USING FUZZY C-MEANS AND K-MEANS METHODS," pp. 1–20, 2022, doi: 10.28919/cmbn/7335.
- [7] A. Ghosal, A. Nandy, A. K. Das, S. Goswami, and M. Panday, "A Short Review on Different Clustering Techniques and Their Applications," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2020, pp. 69–83. doi: 10.1007/978-981-13-7403-6_9.
- [8] C. Muehlethaler and R. Albert, "Collecting data on textiles from the internet using web crawling and web scraping tools," *Forensic Sci Int*, vol. 322, p. 110753, 2021, doi: 10.1016/j.forsciint.2021.110753.
- [9] J. R. Beattie and F. W. L. Esmonde-White, "Exploration of Principal Component Analysis: Deriving Principal Component Analysis Visually Using Spectra," *Appl Spectrosc*, vol. 75, no. 4, pp. 361–375, 2021, doi: 10.1177/0003702820987847.
- [10] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J R Stat Soc Series B Stat Methodol*, vol. 61, no. 2, pp. 611–622, 2002, doi: 10.1111/1467-9868.00196.

- [11] M. Mahmudur, R. Khan, M. Abu Bakr Siddique, R. B. Arif, and M. R. Oishe, "ADBSCAN: Adaptive Density-Based Spatial Clustering of Applications with Noise for Identifying Clusters with Varying Densities." *Energy*, vol. 231, pp. 331–342, Dec. 2018, doi: 10.1016/j.apenergy.2018.09.050.
- [12] Y. Yang *et al.*, "An efficient DBSCAN optimized by arithmetic optimization algorithm with opposition-based learning," *Journal of Supercomputing*, vol. 78, no. 18, pp. 19566–19604, Dec. 2022, doi: 10.1007/s11227-022-04634-w.
- [13] K. Li, Z. Ma, D. Robinson, and J. Ma, "Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering," *Appl*
- [14] E. M. S. ROCHMAN *et al.*, "A combination of algorithm agglomerative hierarchical cluster (AHC) and K-means for clustering tourism in Madura-Indonesia," *Journal of Mathematical and Computational Science*, 2022, doi: 10.28919/jmcs/7086.
- [15] M. Naga Surya Lakshmi and Y. Radhika, "A comparative paper on measuring the performance of snort and suricata with variable packet sizes and speed," *Progress in Color, Colorants and Coatings*, vol. 8, no. 1, pp. 29–37, 2018, doi: 10.14419/ijet.v8i1.13971.