

Klasterisasi Fasilitas Kesehatan BPJS di Indonesia Menggunakan Metode K-Means Clustering

Sandria Amelia Putri¹, Ahmad Wafi Fathurrahman², Alya Setya Paramita³, Mohamad Ibnu Fajar Maulana⁴

Program Studi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur

¹21083010005@student.upnjatim.ac.id, ²21083010011@student.upnjatim.ac.id, ³21083010046@student.upnjatim.ac.id,

⁴21083010106@student.upnjatim.ac.id

Corresponding author email: 21083010005@student.upnjatim.ac.id

Abstract: Basically, there are already many healthcare facilities in Indonesia. However, these facilities are still unevenly distributed across regions. Therefore, this research aims to perform clustering of BPJS healthcare facilities in Indonesia using the K-Means Clustering method and analyze the clustering results. Data on BPJS healthcare facilities, including the quantity and types of facilities in each region of Indonesia, were collected and processed. The K-Means Clustering method was employed to group healthcare facilities based on the number of facilities in each area. The clustering results revealed a clear pattern in the distribution of BPJS healthcare facilities in Indonesia. The Java island region exhibited the highest concentration of healthcare facilities, while other islands had a lower number of facilities. This indicates an imbalance in healthcare facility availability in Indonesia, with Java island being the top priority. This information can serve as a reference for BPJS Kesehatan and relevant parties to improve and develop healthcare services in these regions, ensuring that people throughout Indonesia have equal access to healthcare services.

Keywords: Healthcare facilities, Clustering, K-Means Clustering, BPJS

Abstrak: Pada dasarnya sudah terdapat banyak fasilitas kesehatan yang ada di Indonesia. Namun, fasilitas tersebut masih belum merata di setiap wilayahnya. Maka penelitian ini ditujukan untuk melakukan klasterisasi fasilitas kesehatan BPJS di Indonesia menggunakan metode K-Means Clustering dan menganalisis hasil klasterisasi tersebut. Data fasilitas kesehatan BPJS yang terdiri dari jumlah dan jenis fasilitas kesehatan di setiap wilayah di Indonesia dikumpulkan dan diproses. Metode K-Means Clustering digunakan untuk mengelompokkan fasilitas kesehatan berdasarkan jumlah fasilitas kesehatan di setiap daerah. Hasil klasterisasi menunjukkan bahwa terdapat pola yang jelas dalam distribusi fasilitas kesehatan BPJS di Indonesia. Wilayah pulau Jawa menunjukkan konsentrasi tertinggi fasilitas kesehatan, sedangkan pulau-pulau lainnya memiliki jumlah fasilitas kesehatan yang lebih rendah. Hal ini menunjukkan adanya ketimpangan dalam ketersediaan fasilitas kesehatan di Indonesia, dengan wilayah pulau Jawa menjadi prioritas utama. Hal ini dapat menjadi acuan bagi BPJS Kesehatan dan pihak terkait dalam melakukan perbaikan dan pengembangan pelayanan kesehatan di wilayah-wilayah tersebut, untuk memastikan bahwa masyarakat di seluruh Indonesia mendapatkan akses dan pelayanan kesehatan yang setara.

Kata kunci: Fasilitas kesehatan, Klasterisasi, K-Means Clustering, BPJS

I. PENDAHULUAN

1.1 Latar Belakang

BPJS Kesehatan, sebagai Sistem Jaminan Sosial Kesehatan Nasional, telah diperkenalkan dengan tujuan memberikan akses pelayanan kesehatan yang terjangkau bagi seluruh masyarakat Indonesia. Namun, untuk meningkatkan efisiensi layanan tersebut, BPJS Kesehatan perlu melakukan klasterisasi fasilitas kesehatan yang tersebar di seluruh Indonesia. Salah satu metode yang dapat digunakan untuk melakukan klasterisasi ini adalah metode K-means clustering. Melalui metode K-means clustering, fasilitas kesehatan dapat dikelompokkan berdasarkan jumlah dan jenis fasilitas yang tersedia di setiap daerah. Klasterisasi ini memungkinkan pengelompokan fasilitas kesehatan dengan karakteristik dan profil pelayanan yang serupa. Hasil klasterisasi ini memberikan wawasan berharga dalam pengambilan keputusan terkait alokasi sumber daya, pengembangan layanan kesehatan yang lebih efektif, serta peningkatan koordinasi antara fasilitas kesehatan.

Dengan adanya klasterisasi fasilitas kesehatan, diharapkan BPJS Kesehatan dapat meningkatkan koordinasi antara fasilitas kesehatan dan memperbaiki akses serta kualitas pelayanan kesehatan di seluruh Indonesia. Informasi yang dihasilkan dari klasterisasi ini juga menjadi landasan yang penting dalam pengembangan program dan kebijakan yang sesuai dengan karakteristik masing-masing klaster fasilitas kesehatan. Penelitian mengenai klasterisasi fasilitas kesehatan BPJS Kesehatan dengan menggunakan metode K-means clustering memiliki signifikansi yang besar dalam meningkatkan efisiensi layanan dan memberikan manfaat optimal bagi masyarakat Indonesia. Dengan pemahaman yang lebih baik tentang kelompok fasilitas kesehatan yang ada, BPJS Kesehatan dapat mengoptimalkan sumber daya yang dimilikinya, mengarahkan bantuan dan dukungan yang sesuai, serta merumuskan kebijakan yang tepat untuk meningkatkan pelayanan kesehatan secara keseluruhan.

1.2 Permasalahan

1. Kurangnya kebijakan pemerintah dalam menangani masalah fasilitas kesehatan BPJS di Indonesia
2. Kurangnya klasterisasi fasilitas kesehatan yang mengakibatkan ketidakmerataan adanya fasilitas kesehatan BPJS di Indonesia

1.3 Tujuan

1. Mempermudah pemerintah dalam membuat kebijakan untuk memperluas jangkauan fasilitas kesehatan BPJS di seluruh Indonesia.
2. Menyediakan analisis klasterisasi dalam mempermudah upaya peningkatan fasilitas kesehatan BPJS di Indonesia

1.4 Manfaat

1. Manfaat bagi penulis dan pengetahuan

Klasterisasi dengan algoritma K-Means dalam penelitian ini membantu dalam memahami pola dan karakteristik fasilitas kesehatan BPJS di Indonesia. Hal ini membantu BPJS Kesehatan dan pihak terkait dalam mengidentifikasi perbedaan dan kesamaan antara fasilitas kesehatan tersebut. Klasterisasi membantu meningkatkan koordinasi antara fasilitas kesehatan yang serupa dalam satu klaster. Dengan adanya pertukaran informasi, pengalaman, dan praktik terbaik antar fasilitas kesehatan, kolaborasi dan sinergi dalam memberikan pelayanan kesehatan berkualitas dapat ditingkatkan. Penelitian ini juga akan membantu manusia untuk memperkecil kesalahan dalam melakukan evaluasi terhadap dua model klasterisasi, yaitu K-means Clustering dan Agglomerative Clustering, menggunakan metode silhouette score karena evaluasi ini memberikan pemahaman tentang sejauh mana model klasterisasi yang digunakan dapat menghasilkan klaster yang lebih baik dan valid secara statistik.

2. Manfaat bagi masyarakat

Klasterisasi fasilitas kesehatan BPJS di Indonesia menggunakan algoritma K-Means Clustering akan membantu meningkatkan akses pelayanan kesehatan bagi masyarakat. Dengan memahami pola dan karakteristik fasilitas kesehatan dalam setiap klaster, maka pemerintah juga akan lebih mudah membuat kebijakan fasilitas kesehatan BPJS di Indonesia sehingga klasterisasi ini juga akan memberikan informasi yang lebih akurat dan transparan mengenai fasilitas kesehatan BPJS di Indonesia. Informasi ini dapat diakses oleh masyarakat, sehingga mereka dapat membuat keputusan yang lebih baik terkait pemilihan fasilitas kesehatan yang sesuai dengan kebutuhan dan preferensi mereka.

II. DASAR TEORI

2.1 Badan Penyelenggara Jaminan Nasional (BPJS)

Badan Penyelenggara Jaminan Nasional (BPJS) merupakan

2.2 Klasterisasi

Klasterisasi merupakan

2.3 Algoritma K-Means

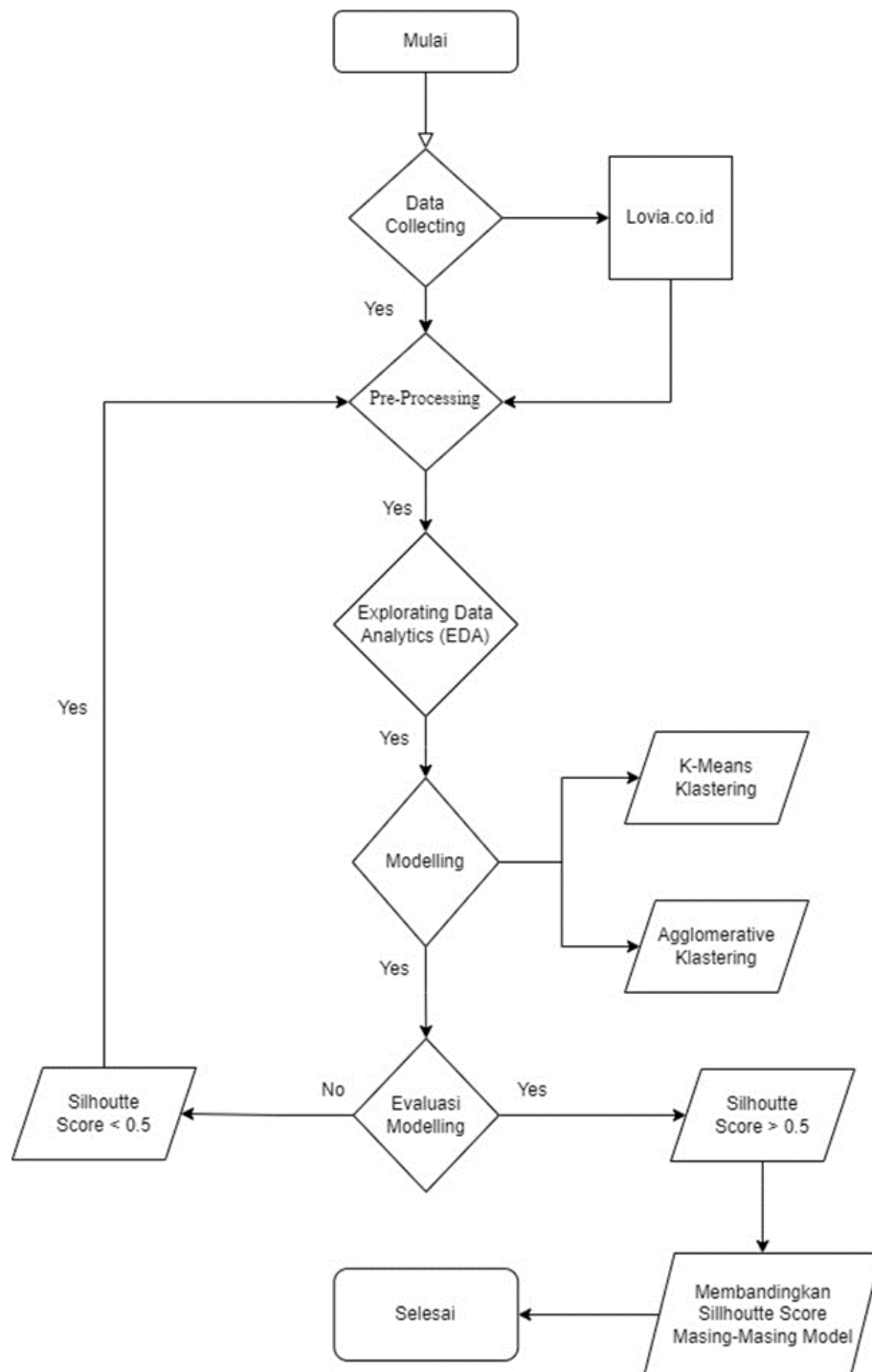
Algoritma k-means merupakan

2.4 Algoritma Agglomerative Hierarchical

Algoritma agglomerative merupakan

III. METODE PENELITIAN

Penelitian ini menggunakan metode klustering dengan algoritma K-Means. Terdapat 5 tahapan dalam penelitian ini, yaitu tahap pengumpulan data (Data Collecting), tahap pengolahan data (Pre-Processing), tahap pemodelan data (Modelling), evaluasi modelling (Silhouette Score). Di dalam prosesnya ingin melakukan klasterisasi list fasilitas kesehatan BPJS di Indonesia yang kemudian mendapatkan output perbandingan dua model, yakni K-means Clustering dengan Agglomerative Clustering dengan melihat *silhouette score*. Adapun diagram alir pada proses penelitian sebagai berikut:



Gambar 1. Flowchart metode penelitian

3.1 Data Collecting

Pada tahap ini dilakukan pengumpulan data yang akan digunakan dalam penelitian. Pada penelitian ini, pengumpulan data dilakukan untuk memperoleh informasi yang diperlukan dalam klasterisasi fasilitas kesehatan BPJS di Indonesia menggunakan metode K-Means Clustering. Data diperoleh dari situs website lovia.co.id yang dilakukan web scraping oleh data engineering kemudian terbentuklah dataset.

3.2 Pre-Processing

Pre-Processing data melibatkan langkah – langkah seperti menangani nilai missing values pada data, penghapusan data yang tidak lengkap atau tidak valid, menangani outliers, menghapus kolom yang tidak perlu, menambahkan kolom yang dibutuhkan, normalisasi data, dan transformasi data lainnya. Pre-Processing data dilakukan untuk meminimalisir adanya bias, meningkatkan interpretasi hasil, dan memungkinkan pengambilan keputusan yang lebih baik.

3.3 Modelling

Pemodelan data dilakukan dengan menggunakan algoritma K-Means. Pada penelitian ini, algoritma K-Means digunakan untuk melakukan klasterisasi fasilitas kesehatan BPJS di Indonesia. Kemudian pada proses ini dilakukan K-Means Clustering serta membandingkan dengan model lainnya seperti Agglomerative Clustering untuk melihat persebaran fasilitas kesehatan BPJS yang tertinggi dan terendah dalam bentuk dua klaster yang selanjutnya melihat perbandingan klasterisasi pada kedua model tersebut.

3.4 Evaluasi Modelling

Proses evaluasi modelling dapat dikatakan baik jika mendapatkan silhouette score akan dikatakan ideal namun, jika mendapatkan silhouette score model model klasterisasi tidak ideal untuk dilakukan. Pada saat melakukan model K-Means Clustering dan Agglomerative Clustering, evaluasi modelling dilakukan dengan metode silhouette score yang akan menampilkan sebuah nilai dari pengukuran dalam menggambarkan seberapa baik objek dalam suatu klaster cocok dengan klaster tersebut. Dengan membandingkan silhouette score K-Means Clustering dan Agglomerative Clustering untuk melihat kualitas klastering yang paling baik.

IV. HASIL DAN PEMBAHASAN

4.1 Data Collecting

Pada tahap awal proses analisis data, pengumpulan data memainkan peran yang penting dalam menghasilkan informasi yang diperlukan untuk penelitian. Data yang kami kumpulkan diambil dari website lovia yang telah dilakukan web scraping oleh seorang data engineer. Lovia merupakan suatu website konseling profesional yang di dalamnya terdapat beberapa sub bagian, salah satunya mengenai kesehatan. Dataset tersebut berupa list fasilitas kesehatan (faskes) yang bekerja sama dengan BPJS.

NoLink	Provinsi	KotaKab	Link	TipeFaskes	No	KodeFaskes	NamaFaskes	LatLongFaskes	AlamatFaskes	TelpFaskes
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Rumah Sakit	1	0001R001	RSU Cut Nyak Dhien ...	http://maps.google.co...	Jl. Tm Bahrum No. 1 ...	0641-0621039
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Rumah Sakit	2	0001R004	RSU Cut Meutia Lang...	http://maps.google.co...	Jl. Garuda Kebun Bar...	0641-4840076
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Rumah Sakit	3	0105R001	RSUD Langsa ...	http://maps.google.co...	Jln. A. Yani No. 1 Langsa	0641-22051
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Rumah Sakit	4	0105R013	RSU Ummi ...	http://maps.google.co...	Jln. Prof. A. Madjid Ib...	0641-22886
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Puskesmas	1	00010001	Langsa Lama ...	http://maps.google.co...	Gampong Meurandeh...	0641-064121218
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Puskesmas	2	00010801	Langsa Baro ...	http://maps.google.co...	Jl. Liliawangsa Langsa	0000-0000000
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Puskesmas	3	00010802	Langsa Barat ...	http://maps.google.co...	Jl. Prof. A. Madjid Ibra...	0000-064120282
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Puskesmas	4	00010803	Langsa Kota ...	http://maps.google.co...	Jl. H. Agussalim No.10	0641-064121912
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Puskesmas	5	00011801	Langsa Timur ...	http://maps.google.co...	Medan-Banda Aceh K...	0821-65824360
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Dokter Praktik Perora...	-	-	-	-	-	-
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Dokter Gigi	-	-	-	-	-	-
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Klinik Utama	1	0001U011	dr. Hj. Sri Hariaty ...	http://maps.google.co...	Jalan Panglima Pole...	0811-6708681
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Klinik Utama	2	0001U013	dr. Herman. I ...	http://maps.google.co...	Medan-Banda Aceh, ...	0811-6700396
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Klinik Utama	3	0001U014	dr. MUHAMMAD YUS...	http://maps.google.co...	Medan - Banda Aceh ...	0811-6772220
0	Nanggroe Aceh Daru...	Kode Faskes dan Ala...	https://ovia.lfe/d/hea...	Klinik Utama	4	0001U015	Dr. Redha Dian Akbar...	-	Liliawangsa, Geduba...	0813-70059510

Gambar 2. Dataset

4.2 Data Cleaning

Data cleaning dilakukan untuk memastikan kebersihan dan kualitas data yang digunakan dalam analisis. Kami telah melakukan serangkaian data cleaning yang meliputi langkah-langkah berikut:

4.2.1 Menghapus spasi dan kolom yang tidak diperlukan

Pada kolom ‘NamaFaskes’ dan ‘TelpFaskes’ terdapat spasi yang berlebih sehingga perlu dilakukan penghapusan spasi tersebut. Selanjutnya, kolom ‘Link’, ‘NoLink’, dan ‘No’ kami rasa tidak diperlukan sehingga kolom-kolom tersebut perlu dihapus. Berikut merupakan dataset setelah dilakukan beberapa penghapusan.

	Provinsi	KotaKab	TipeFaskes	KodeFaskes	NamaFaskes	LatLongFaskes	AlamatFaskes	TelpFaskes
0	Nanggroe Aceh Darussalam	Kota Langsa	Rumah Sakit	0001R001	RSU Cut Nyak Dhien	http://maps.google.co.id/?q=4.488058,97.947963	Jl. Tm Bahrum No. 1 Langsa	0641-0621039
1	Nanggroe Aceh Darussalam	Kota Langsa	Rumah Sakit	0001R004	RSU Cut Meutia Langsa	http://maps.google.co.id/?q=4.488088,97.947781	Jl. Garuda Kebun Baru Langsa	0641-4840076
2	Nanggroe Aceh Darussalam	Kota Langsa	Rumah Sakit	0105R001	RSUD Langsa	http://maps.google.co.id/?q=4.472208,97.975533	Jln. A. Yani No. 1 Langsa	0641-22051
3	Nanggroe Aceh Darussalam	Kota Langsa	Rumah Sakit	0105R013	RSU Ummi	http://maps.google.co.id/?q=4.470376,97.991915	Jln. Prof. A. Madjid Ibrahim	0641-22886
4	Nanggroe Aceh Darussalam	Kota Langsa	Puskesmas	00010001	Langsa Lama	http://maps.google.co.id/?q=4.478172,97.949988	Gampong Meurandeh Dayah	0641-064121218

Gambar 3. Dataset setelah dilakukan beberapa penghapusan

4.2.2 Memproses data latitude dan longitude

Data koordinat latitude dan longitude yang diberikan berupa format link yang membuatnya sulit dipahami. Untuk memudahkan pemahaman dan keterbacaan, kami akan melakukan pemrosesan pada data tersebut untuk mengubahnya menjadi titik koordinat yang lebih akurat. Berikut merupakan dataset setelah dilakukan pemrosesan data latitude dan longitude.

	Provinsi	KotaKab	TipeFaskes	KodeFaskes	NamaFaskes	AlamatFaskes	TelpFaskes	Latitude	Longitude
0	Nanggroe Aceh Darussalam	Kota Langsa	Rumah Sakit	0001R001	RSU Cut Nyak Dhien	Jl. Tm Bahrum No. 1 Langsa	0641-0621039	4.488058	97.947963
1	Nanggroe Aceh Darussalam	Kota Langsa	Rumah Sakit	0001R004	RSU Cut Meutia Langsa	Jl.Garuda Kebun Baru Langsa	0641-4840076	4.488088	97.947781
2	Nanggroe Aceh Darussalam	Kota Langsa	Rumah Sakit	0105R001	RSUD Langsa	Jln.A.Yani No. 1 Langsa	0641-22051	4.472208	97.975533
3	Nanggroe Aceh Darussalam	Kota Langsa	Rumah Sakit	0105R013	RSU Ummi	Jln. Prof. A. Madjid Ibrahim	0641-22886	4.470376	97.991915
4	Nanggroe Aceh Darussalam	Kota Langsa	Puskesmas	00010001	Langsa Lama	Gampong Meurandeh Dayah	0641-064121218	4.478172	97.949988

Gambar 4. Dataset setelah dilakukan pemrosesan latitude dan longitude

4.2.3 Memeriksa missing value dan baris redundant

Pada tahap ini, kami melakukan identifikasi dan penanganan nilai yang hilang serta deteksi dan penghapusan baris yang redundan dalam dataset. Namun, didapatkan dari dataset list faskes BPJS di Indonesia tidak ditemukan missing value dan baris yang redundan. Sehingga kebersihan dan kualitas data yang digunakan dalam analisis sudah dapat dipastikan serta data sudah menjadi lebih akurat dan relevan untuk analisis yang lebih baik.

```

Provinsi tidak terdapat missing value!
KotaKab tidak terdapat missing value!
TipeFaskes tidak terdapat missing value!
KodeFaskes tidak terdapat missing value!
NamaFaskes tidak terdapat missing value!
LatLongFaskes tidak terdapat missing value!
AlamatFaskes tidak terdapat missing value!
TelpFaskes tidak terdapat missing value!
Data tidak memiliki baris redundant!

```

Gambar 5. Hasil pengecekan missing value dan baris redundant

4.3 EDA (Exploratory Data Analysis)

EDA dilakukan untuk menggali dan memahami karakteristik, pola, dan hubungan dalam dataset secara visual dan deskriptif. Tujuan utama dari EDA adalah untuk mendapatkan wawasan awal tentang data sebelum dilakukan analisis statistik yang lebih mendalam atau pengembangan model.. Kami telah melakukan serangkaian EDA yang meliputi langkah-langkah berikut:

4.3.1 Info dataset

Info dataset yang kami lakukan menunjukkan informasi tentang struktur dataset yang terdiri dari 9 kolom dan 29.157 baris. Beberapa hal penting yang diperoleh yaitu jumlah data, tipe data, nilai yang hilang, dan penggunaan memori. Berikut merupakan info dari dataset list faskes BPJS di Indonesia.

```

Info dataset:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 29157 entries, 0 to 29156
Data columns (total 9 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   Provinsi            29157 non-null  object  
1   KotaKab             29157 non-null  object  
2   TipeFaskes          29157 non-null  object  
3   KodeFaskes          29157 non-null  object  
4   NamaFaskes          29157 non-null  object  
5   AlamatFaskes        29157 non-null  object  
6   TelpFaskes          29157 non-null  object  
7   Latitude            16517 non-null  float64 
8   Longitude           19729 non-null  float64 
dtypes: float64(2), object(7)
memory usage: 2.0+ MB

```

Gambar 6. Info dataset

4.3.2 Menampilkan jumlah unik untuk kolom kategorisasi

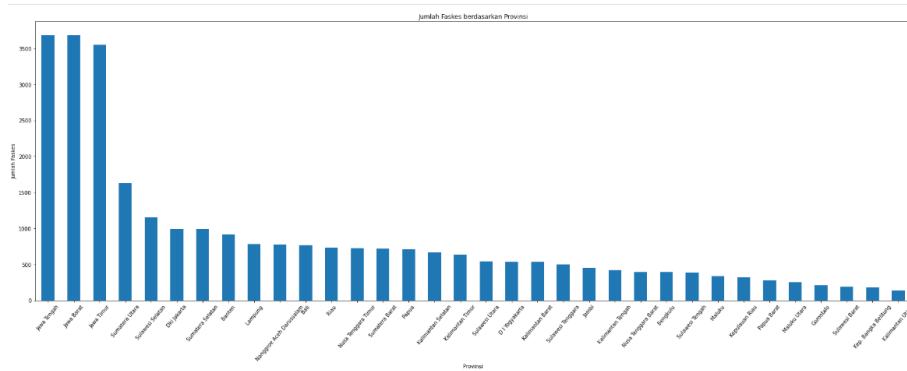
Mengetahui jumlah unik dari kolom kategorisasi membantu kita dalam memahami variasi atau keragaman data dalam setiap kategori. Ini penting untuk memahami sebaran data dan mengidentifikasi kategori yang mungkin lebih dominan atau kurang mewakili. Jumlah unik juga dapat memberikan gambaran tentang distribusi data dan memberikan informasi tentang apakah ada kategori yang sangat dominan atau jarang terjadi dalam dataset. Berikut merupakan jumlah unik untuk kolom kategorisasi dataset list faskes BPJS di Indonesia.

Jumlah unik Provinsi: 34		
Jawa Tengah	3688	
Jawa Barat	3683	
Jawa Timur	3548	
Sumatera Utara	1628	
Sulawesi Selatan	1155	
Dki Jakarta	995	
Sumatera Selatan	993	
Banten	913	
Lampung	784	
Nanggroe Aceh Darussalam	776	
Bali	765	
Riau	729	
Nusa Tenggara Timur	722	
Sumatera Barat	721	
Papua	712	
Kalimantan Selatan	671	
Kalimantan Timur	634	
Sulawesi Utara	543	
D I Yogyakarta	534	
Kalimantan Barat	531	
Sulawesi Tenggara	500	
Jambi	446	
Kalimantan Tengah	418	
Nusa Tenggara Barat	393	
Bengkulu	390	
Sulawesi Tengah	386	
Maluku	334	
Kepulauan Riau	319	
Papua Barat	278	
Maluku Utara	253	
Gorontalo	206	
Sulawesi Barat	187	
Kep. Bangka Belitung	182	
Kalimantan Utara	140	
Name: Provinsi, dtype: int64		
Jumlah unik KotaKab: 512		
Kota Surabaya	364	
Kota Semarang	305	
Kab. Bogor	298	
Kota Makassar	292	
Kota Palembang	288	
Kab.Bolaang Mongondow Tmr	12	
Kab. Banggai Laut	12	
Kab. Mamberamo Tengah	11	
Kab. Tana Tidung	11	
Kab. Manokwari Selatan	10	
Name: KotaKab, Length: 512, dtype: int64		
Jumlah unik TipeFaskes: 7		
Puskesmas	9913	
Klinik Pratama	6597	
Klinik Utama	5375	
Apotek	2982	
Rumah Sakit	2249	
Dokter Gigi	1410	
Dokter Praktik Perorangan	631	
Name: TipeFaskes, dtype: int64		

Gambar 7. Jumlah unik setiap kolom kategorisasi

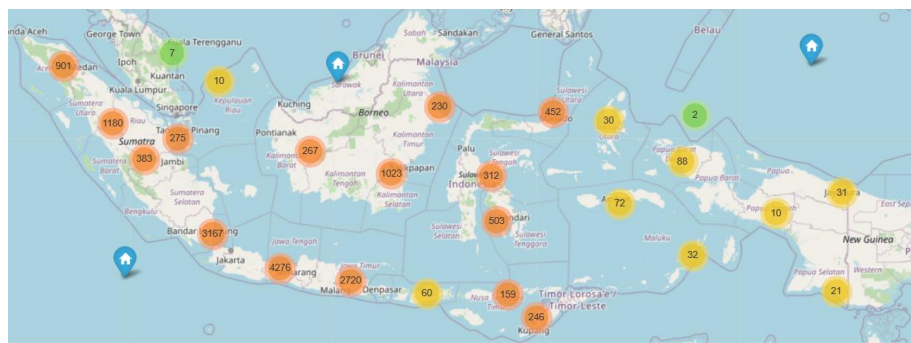
4.3.3 Visualisasi jumlah faskes berdasarkan provinsi

Dengan melihat visualisasi jumlah faskes berdasarkan provinsi, terlihat bahwa Provinsi Jawa Tengah memiliki jumlah faskes yang paling tinggi dan Provinsi Kalimantan Utara memiliki jumlah faskes yang paling rendah. Berikut merupakan hasil visualisasinya.



4.3.4 Visualisasi geografis faskes BPJS di Indonesia

Dengan menggunakan visualisasi geografis, kita dapat dengan jelas melihat sebaran fasilitas kesehatan BPJS di seluruh wilayah Indonesia. Visualisasi ini memberikan gambaran yang komprehensif tentang distribusi fasilitas kesehatan dengan menggunakan kode warna. Wilayah dengan angka berwarna oranye menunjukkan jumlah fasilitas kesehatan yang tinggi, yaitu lebih dari 150. Wilayah dengan angka berwarna kuning menunjukkan jumlah fasilitas kesehatan yang cukup, yaitu antara 10 hingga 90. Terakhir, wilayah dengan angka berwarna hijau menunjukkan jumlah fasilitas kesehatan yang rendah, yaitu kurang dari 10. Berikut merupakan hasil visualisasinya.



4.3.5 Menghitung jumlah faskes yang dapat divisualisasikan

Dalam dataset yang diberikan, terdapat beberapa entri yang berada di luar batas geografis Indonesia. Maka dari itu, kami melakukan pemfilteran data fasilitas kesehatan berdasarkan batas geografis Indonesia, membuat peta interaktif dengan menampilkan titik-titik fasilitas kesehatan yang dapat ditampilkan di peta, dan memberikan informasi tentang jumlah total data fasilitas kesehatan serta jumlah fasilitas yang dapat dan tidak dapat ditampilkan di peta. Berikut merupakan hasil perhitungannya.

Total data fasilitas: 29157

Total fasilitas yang dapat ditampilkan di map: 16460

Total fasilitas yang tidak dapat ditampilkan di map: 12697

Gambar 10. Perhitungan jumlah faskes yang dapat divisualisasikan

4.3.6 Menghitung jumlah faskes di setiap kota/kabupaten

Dengan menghitung jumlah faskes di setiap kota/kabupaten, terlihat 5 kota teratas yang memiliki jumlah faskes tertinggi, di antaranya yaitu Kota Surabaya, Kota Semarang, Kab. Bogor, Kota Makassar, dan Kota Palembang. Berikut merupakan hasil perhitungannya.

	KotaKab	Jumlah
0	Kota Surabaya	364
1	Kota Semarang	305
2	Kab. Bogor	298
3	Kota Makassar	292
4	Kota Palembang	288

Gambar 11. Perhitungan jumlah faskes di setiap kota/kabupaten

4.4 Clustering

Kami melakukan analisis klastering pada dataset yang berisi daftar fasilitas kesehatan BPJS di Indonesia dengan tujuan untuk mengidentifikasi klaster tertinggi dan terendah. Dalam analisis ini, kami menggunakan dua metode klastering, yaitu k-means dan agglomerative clustering. Berikut merupakan penjelasannya:

4.4.1 K-Means clustering

Kami menggunakan metode K-Means clustering dengan dua klaster untuk menganalisis data fasilitas kesehatan BPJS di Indonesia. Tujuan kami adalah mengidentifikasi klaster tertinggi dan terendah berdasarkan jumlah faskes. Dengan melakukan klasterisasi, kami dapat mengelompokkan fasilitas kesehatan ke dalam dua kelompok yang memiliki karakteristik yang serupa. Klaster tertinggi ditandai dengan warna merah, sementara klaster terendah ditandai dengan warna biru. Hal ini membantu kami memahami pola dan perbedaan antara klaster tertinggi dan terendah, sehingga memberikan wawasan yang berharga dalam pengambilan keputusan terkait penyediaan pelayanan kesehatan di Indonesia.

```
# Membuat peta dengan pewarnaan berdasarkan klaster
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
import folium

# Mengambil data jumlah faskes per kota/kabupaten
data_cluster = merged[['Jumlah']]

# Melakukan klastering dengan K-Means
kmeans = KMeans(n_clusters=2, random_state=0)
kmeans.fit(data_cluster)
```

```
# Menambahkan label klaster ke data_cluster
data_cluster['Cluster'] = kmeans.labels_

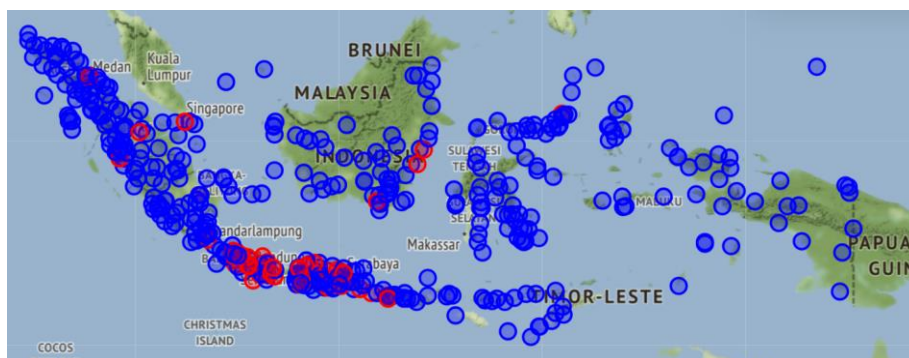
cluster_map = folium.Map(location=[data_n['Latitude'].mean(),
data_n['Longitude'].mean()], zoom_start=6, tiles='Stamen
Terrain')

colors = ['red', 'blue']

for lat, lon, prov, kotakab, jml, cluster in
zip(merged['Latitude'], merged['Longitude'],
merged['Provinsi'], merged['KotaKab'], merged['Jumlah'],
data_map['Cluster']):
    folium.CircleMarker([lat, lon],
                        radius=10,
                        popup = ('<strong>Provinsi</strong>:
' + str(prov) + '<br>'
                                '<strong>Kota/Kab</strong>: '
+ str(kotakab) + '<br>'
                                '<strong>Jumlah</strong>: ' +
str(jml) + '<br>'
                                '<strong>Cluster</strong>: '
+ str(cluster)),
                        color=colors[cluster],
                        fill_color=colors[cluster],
                        fill_opacity=0.3).add_to(cluster_map)
cluster_map
```

Gambar 12. Potongan script K-Means clustering

Dari hasil visualisasi geografis, terlihat bahwa terdapat banyak titik berwarna merah pada Pulau Jawa. Hal ini berarti bahwa Pulau Jawa memiliki faskes BPJS yang tinggi. Berikut merupakan visualisasinya.



Gambar 13. Visualisasi geografis K-Means clustering

4.4.2 Agglomerative clustering

Kami menggunakan metode Agglomerative Clustering untuk menganalisis data fasilitas kesehatan BPJS di Indonesia dengan tujuan mengidentifikasi klaster tertinggi dan

terendah berdasarkan jumlah faskes. Dalam metode ini, kami mengelompokkan fasilitas kesehatan ke dalam dua kelompok yang memiliki karakteristik serupa. Kluster tertinggi ditandai dengan warna merah, sedangkan kluster terendah ditandai dengan warna biru. Melalui analisis kluster ini, kami dapat memperoleh pemahaman yang mendalam mengenai pola dan perbedaan antara kluster tertinggi dan terendah, memberikan wawasan berharga dalam pengambilan keputusan terkait penyediaan pelayanan kesehatan di Indonesia.

```
from sklearn.cluster import AgglomerativeClustering
from sklearn.metrics import silhouette_score

# Mengambil data jumlah faskes per kota/kabupaten
data_cluster = merged[['Jumlah']]

# Melakukan klastering dengan Agglomerative Clustering
n_clusters = 2 # Jumlah kluster yang diinginkan
agglomerative = AgglomerativeClustering(n_clusters=n_clusters)
agglomerative.fit(data_cluster)

# Menambahkan label kluster ke data_cluster
data_cluster['Cluster'] = agglomerative.labels_

# Visualisasi peta dengan pewarnaan berdasarkan kluster
cluster_map = folium.Map(location=[data_n['Latitude'].mean(),
                                     data_n['Longitude'].mean()], zoom_start=6, tiles='Stamen Terrain')

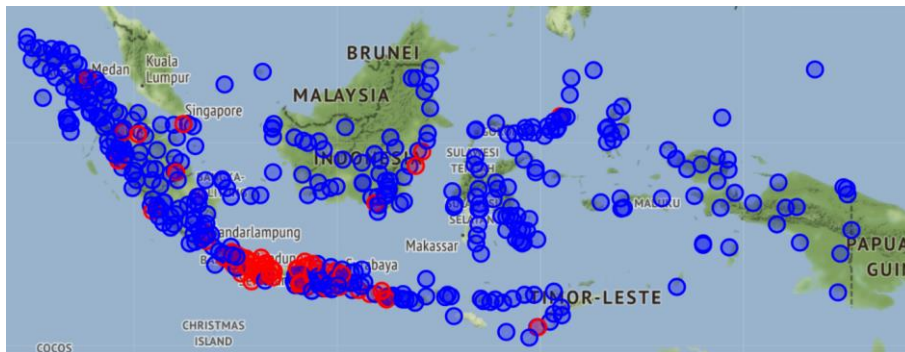
colors = ['red', 'blue']

for lat, lon, prov, kotakab, jml, cluster in zip(merged['Latitude'], merged['Langitude'],
                                                merged['Provinsi'], merged['KotaKab'], merged['Jumlah'],
                                                data_cluster['Cluster']):
    folium.CircleMarker([lat, lon], radius=10, popup = ('<strong>Provinsi</strong>:' + str(prov) + '<br>' + str(kotakab) + '<br>' + str(jml) + '<br>' + str(cluster)), color=colors[cluster], fill_color=colors[cluster], fill_opacity=0.3).add_to(cluster map)
```

```
cluster_map
```

Gambar 14. Potongan script Agglomerative clustering

Sama seperti hasil visualisasi geografis K-Means clustering, pada visualisasi geografis agglomerative clustering juga terlihat bahwa terdapat banyak titik berwarna merah pada Pulau Jawa. Hal ini berarti bahwa Pulau Jawa memiliki faskes BPJS yang tinggi. Berikut merupakan visualisasinya.



Gambar 15. Visualisasi geografis Agglomerative clustering

4.5 Evaluasi Silhouette Score

Evaluasi Silhouette Score digunakan untuk mengukur sejauh mana kualitas klustering yang telah dilakukan. Silhouette Score memberikan penilaian terhadap seberapa baik objek dalam sebuah klaster berada dalam kelompoknya sendiri dibandingkan dengan kelompok lain. Berikut merupakan perhitungannya:

4.5.1 K-Means clustering

Agar lebih mudah dipahami, kami membuat grafik silhoutte plot untuk K-Means clustering yang menunjukkan nilai silhouette coefficient untuk setiap klaster, dengan daerah berwarna menggambarkan seberapa padatnya anggota klaster tersebut. Garis vertikal berwarna merah menunjukkan nilai silhouette score keseluruhan. Grafik ini membantu kita memahami sejauh mana setiap klaster terpisah satu sama lain, serta seberapa baik kualitas klustering yang telah dilakukan.

```
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score, silhouette_samples
import folium
import matplotlib.pyplot as plt
import numpy as np

# Mengambil data jumlah faskes per kota/kabupaten
```

```

data_cluster = merged[['Jumlah']]

# Melakukan klastering dengan K-Means
kmeans = KMeans(n_clusters=2, random_state=0)
kmeans.fit(data_cluster)

# Menambahkan label klaster ke data_cluster
data_cluster.loc[:, 'Cluster'] = kmeans.labels_

# Menghitung Silhouette Score
silhouette_avg = silhouette_score(data_cluster,
kmeans.labels_)

# Visualisasi peta dengan pewarnaan berdasarkan klaster
cluster_map = folium.Map(location=[data_n['Latitude'].mean(),
data_n['Longitude'].mean()], zoom_start=6, tiles='Stamen
Terrain')

colors = ['red', 'blue']

for lat, lon, prov, kotakab, jml, cluster in
zip(merged['Latitude'], merged['Longitude'],
merged['Provinsi'], merged['KotaKab'], merged['Jumlah'],
data_map['Cluster']):
    folium.CircleMarker([lat, lon],
                        radius=10,
                        popup = ('<strong>Provinsi</strong>:
' + str(prov) + '<br>'
                                '<strong>Kota/Kab</strong>: '
+ str(kotakab) + '<br>'
                                '<strong>Jumlah</strong>: ' +
str(jml) + '<br>'
                                '<strong>Cluster</strong>: '
+ str(cluster)),
                        color=colors[cluster],
                        fill_color=colors[cluster],
                        fill_opacity=0.3).add_to(cluster_map)

# Visualisasi grafik Silhouette Score
fig, ax = plt.subplots(figsize=(6, 4))
cluster_labels = np.unique(kmeans.labels_)
silhouette_values = silhouette_samples(data_cluster,
kmeans.labels_)

y_lower = 10
for i, label in enumerate(cluster_labels):
    cluster_silhouette_values =
silhouette_values[kmeans.labels_ == label]
    cluster_silhouette_values.sort()
    cluster_size = cluster_silhouette_values.shape[0]
    y_upper = y_lower + cluster_size

```

```

        color = colors[label]
        ax.fill_betweenx(np.arange(y_lower, y_upper),
                        0, cluster_silhouette_values,
                        facecolor=color, edgecolor=color,
alpha=0.7)

        ax.text(-0.05, y_lower + 0.5 * cluster_size, str(label))
        y_lower = y_upper + 10

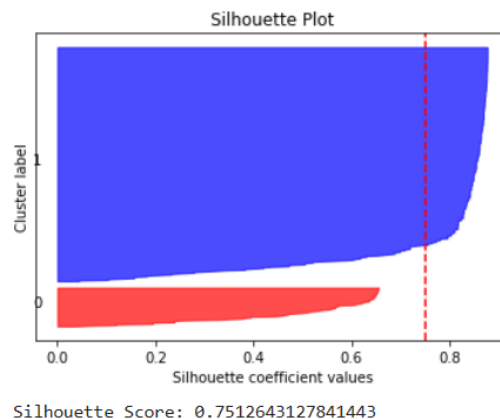
ax.set_xlabel("Silhouette coefficient values")
ax.set_ylabel("Cluster label")
ax.set_title("Silhouette Plot")
ax.axvline(x=silhouette_avg, color="red", linestyle="--")
ax.yaxis.set_ticks([])
plt.show()

print("Silhouette Score:", silhouette_avg)

```

Gambar 16. Potongan script silhouette score K-Means clustering

Nilai silhouette score yang didapatkan untuk metode K-Means yaitu sebesar 0.7512643127841443. Nilai tersebut menunjukkan bahwa klastering yang dilakukan dengan metode K-Means menghasilkan hasil yang baik. Silhouette score berkisar dari -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan bahwa klastering lebih baik. Secara khusus, nilai tersebut menunjukkan bahwa setiap data dalam klaster memiliki jarak yang cukup jauh dengan data di klaster lain. Selain itu, data dalam klaster ini juga memiliki kohesi yang tinggi, artinya mereka cenderung saling berdekatan dalam ruang fitur.



Gambar 17. Silhouette score K-Means clustering

4.5.2 Agglomerative clustering

Agar lebih mudah dipahami, kami membuat grafik silhouette plot untuk Agglomerative clustering yang menunjukkan nilai silhouette coefficient untuk setiap klaster, dengan daerah berwarna menggambarkan seberapa padatnya anggota klaster tersebut. Garis vertikal berwarna merah menunjukkan nilai silhouette score keseluruhan. Grafik ini membantu kita memahami

sejauh mana setiap klaster terpisah satu sama lain, serta seberapa baik kualitas klastering yang telah dilakukan.

```
import pandas as pd
from sklearn.cluster import AgglomerativeClustering
from sklearn.metrics import silhouette_score, silhouette_samples
import folium
import matplotlib.pyplot as plt
import numpy as np

# Mengambil data jumlah faskes per kota/kabupaten
data_cluster = merged[['Jumlah']]

# Melakukan klastering dengan Agglomerative Clustering
n_clusters = 2
agglomerative = AgglomerativeClustering(n_clusters=n_clusters)
agglomerative.fit(data_cluster)

# Menambahkan label klaster ke data_cluster
data_cluster.loc[:, 'Cluster'] = agglomerative.labels_

# Menghitung Silhouette Score
silhouette_avg = silhouette_score(data_cluster, agglomerative.labels_)

# Visualisasi peta dengan pewarnaan berdasarkan klaster
cluster_map = folium.Map(location=[data_n['Latitude'].mean(),
data_n['Longitude'].mean()], zoom_start=6, tiles='Stamen Terrain')

colors = ['red', 'blue']

for lat, lon, prov, kotakab, jml, cluster in zip(merged['Latitude'], merged['Longitude'], merged['Provinsi'], merged['KotaKab'], merged['Jumlah'], data_map['Cluster']):
    folium.CircleMarker([lat, lon],
                        radius=10,
                        popup = ('<strong>Provinsi</strong>: '
+ str(prov) + '<br>'
+ str(kotakab) + '<br>'
+ str(jml) + '<br>'
+ str(cluster)),
                        color=colors[cluster],
                        fill_color=colors[cluster],
```



```

fill_opacity=0.3).add_to(cluster_map)

# Visualisasi grafik Silhouette Score
fig, ax = plt.subplots(figsize=(6, 4))
cluster_labels = np.unique(agglomerative.labels_)
silhouette_values = silhouette_samples(data_cluster,
agglomerative.labels_)

y_lower = 10
for i, label in enumerate(cluster_labels):
    cluster_silhouette_values =
silhouette_values[agglomerative.labels_ == label]
    cluster_silhouette_values.sort()
    cluster_size = cluster_silhouette_values.shape[0]
    y_upper = y_lower + cluster_size

    color = colors[label]
    ax.fill_betweenx(np.arange(y_lower, y_upper),
0, cluster_silhouette_values,
facecolor=color, edgecolor=color,
alpha=0.7)

    ax.text(-0.05, y_lower + 0.5 * cluster_size, str(label))
    y_lower = y_upper + 10

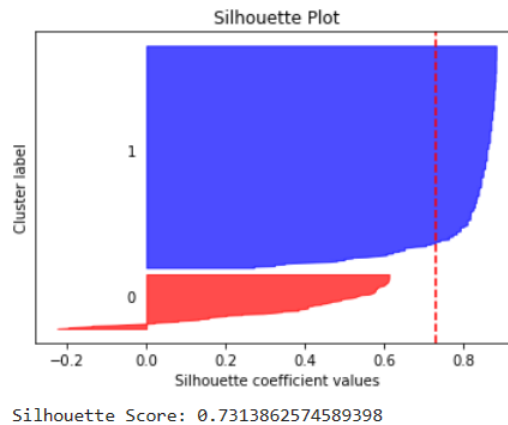
ax.set_xlabel("Silhouette coefficient values")
ax.set_ylabel("Cluster label")
ax.set_title("Silhouette Plot")
ax.axvline(x=silhouette_avg, color="red", linestyle="--")
ax.yaxis.set_ticks([])
plt.show()

print("Silhouette Score:", silhouette_avg)

```

Gambar 18. Potongan script silhouette score Agglomerative clustering

Pada metode Agglomerative Clustering, nilai Silhouette Score sebesar 0.7313862574589398 juga menunjukkan hasil yang baik. Ini mengindikasikan bahwa pembagian fasilitas kesehatan BPJS ke dalam klaster menggunakan metode Agglomerative Clustering memiliki tingkat kohesi yang tinggi di dalam klaster dan pemisahan yang jelas antara klaster.



Gambar 19. Silhouette score Agglomerative clustering

4.5.3 Perbandingan silhouette score

Kami melakukan perbandingan silhouette score antara kedua metode untuk mengetahui bahwa modelling yang dipilih yaitu K-Means clustering merupakan metode paling baik dibandingkan modelling Agglomerative clustering.

```
import matplotlib.pyplot as plt

# K-Means Clustering
kmeans = KMeans(n_clusters=2, random_state=0)
kmeans.fit(data_cluster)
kmeans_score = silhouette_score(data_cluster, kmeans.labels_)

# Agglomerative Clustering
agglomerative = AgglomerativeClustering(n_clusters=2)
agglomerative.fit(data_cluster)
agglomerative_score = silhouette_score(data_cluster,
agglomerative.labels_)

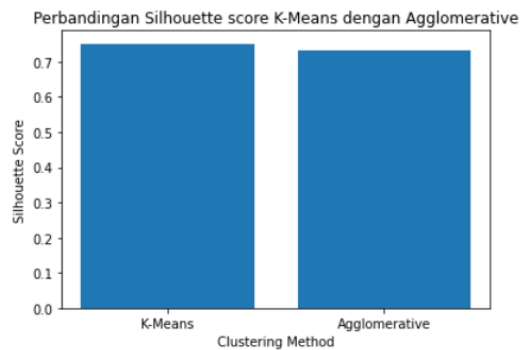
# Menampilkan Silhouette Score
print("Silhouette Score K-Means:", kmeans_score)
print("Silhouette Score Agglomerative Clustering:",
agglomerative_score)

# Membuat grafik perbandingan Silhouette Score
labels = ['K-Means', 'Agglomerative']
scores = [kmeans_score, agglomerative_score]

plt.bar(labels, scores)
plt.xlabel('Clustering Method')
plt.ylabel('Silhouette Score')
plt.title('Perbandingan Silhouette score K-Means dengan
Agglomerative')
plt.show()
```

Gambar 20. Potongan script perbandingan silhouette score

Hasil Silhouette Score pada K-Means clustering menunjukkan keunggulan sebesar 0.01988 dibandingkan dengan metode Agglomerative clustering. Oleh karena itu, dapat disimpulkan bahwa metode K-Means clustering adalah pilihan terbaik untuk melakukan klasterisasi fasilitas kesehatan BPJS di Indonesia.



Gambar 21. Bar plot perbandingan silhouette score

4.6 Hasil Klasterisasi

Selanjutnya, kami akan menampilkan visualisasi 25 kabupaten/kota dengan jumlah faskes tertinggi dalam klaster tertinggi, serta 25 kabupaten/kota dengan jumlah faskes terendah dalam klaster terendah. Visualisasi ini didasarkan pada hasil klasterisasi menggunakan metode K-Means clustering.

```
import pandas as pd
from sklearn.cluster import KMeans
import seaborn as sns
import matplotlib.pyplot as plt

# Preprocessing data
data_kotakab = df['KotaKab'].value_counts().reset_index()
data_kotakab.columns = ['KotaKab', 'Jumlah']

# K-means clustering
features = data_kotakab[['Jumlah']]
kmeans = KMeans(n_clusters=2, random_state=42)
kmeans.fit(features)
labels = kmeans.labels_

# Add Cluster column to the dataframe
data_kotakab['Cluster'] = labels

# Sort data by Jumlah in descending order
data_kotakab = data_kotakab.sort_values(by='Jumlah',
ascending=False)

# Create a temporary dataframe for visualization
```

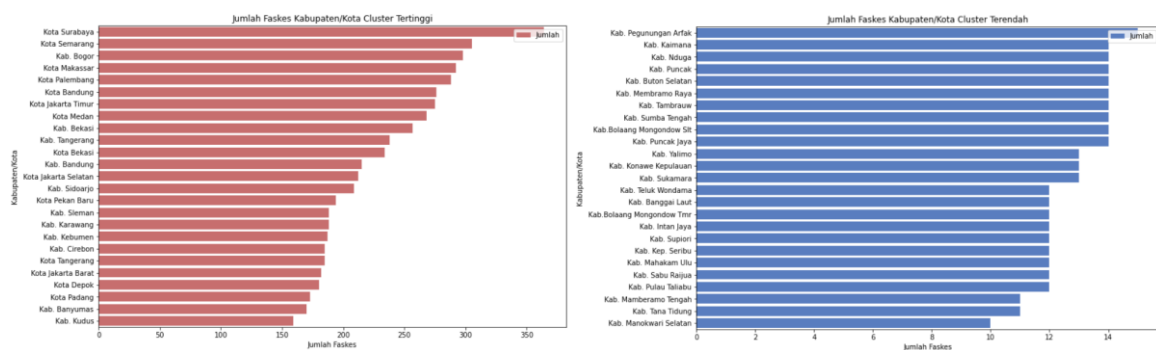
```
temp_data = pd.DataFrame({'KotaKab': data_kotakab['KotaKab'],
                          'Jumlah': data_kotakab['Jumlah']})

# Plotting the bar chart for highest faskes
f, ax = plt.subplots(figsize=(12, 8))
sns.set_color_codes("muted")
sns.barplot(x="Jumlah", y="KotaKab", data=temp_data[:25],
            label="Jumlah", color="r")
plt.xlabel('Jumlah Faskes')
plt.ylabel('Kabupaten/Kota')
plt.title('Jumlah Faskes Kabupaten/Kota Cluster Tertinggi')
plt.legend(loc='upper right')
plt.show()

# Plotting the bar chart for lowest faskes
f, ax = plt.subplots(figsize=(12, 8))
sns.set_color_codes("muted")
sns.barplot(x="Jumlah", y="KotaKab", data=temp_data[-25:],
            label="Jumlah", color="b")
plt.xlabel('Jumlah Faskes')
plt.ylabel('Kabupaten/Kota')
plt.title('Jumlah Faskes Kabupaten/Kota Cluster Terendah')
plt.legend(loc='upper right')
plt.show()
```

Gambar 22. Potongan script hasil klasterisasi

Berikut adalah hasil visualisasi klaster yang diperoleh. Terlihat bahwa Kota Surabaya menduduki peringkat pertama dalam klaster dengan jumlah faskes tertinggi, sedangkan Kabupaten Manokwari Selatan menduduki peringkat terakhir dalam klaster dengan jumlah faskes terendah.



Gambar 23. Hasil klasterisasi

V. KESIMPULAN

Klasterisasi fasilitas kesehatan BPJS di Indonesia menggunakan metode K-Means Clustering menghasilkan dua klaster dengan kualitas yang berbeda. Klaster pertama merupakan klaster fasilitas kesehatan tertinggi, sedangkan klaster kedua adalah klaster fasilitas kesehatan terendah. Distribusi kedua klaster tersebut tergambar pada peta wilayah Indonesia.

Hasil klasterisasi menunjukkan bahwa fasilitas kesehatan tertinggi terdapat di wilayah pulau Jawa, sementara fasilitas kesehatan di pulau-pulau lainnya memiliki kualitas yang lebih rendah. Temuan ini menjadi perhatian penting bagi pemerintah untuk melakukan evaluasi guna memperhatikan dan mengembangkan fasilitas kesehatan di pulau-pulau selain Jawa, sebagai upaya untuk meratakan pelayanan kesehatan di seluruh Indonesia.

3. REFERENSI

<https://tanyadigital.com/data-collection-adalah/> (REFERENSI UNTUK DATA COLLECTING BAB 3)

Sun Binli, "Research on data-preprocessing for construction of university information systems", IEEE Conference. Computer Application and System Modeling (ICCASM), 2010 International Conference Volume:1. 22-24 Oct. 2010 (REFERENSI UNTUK PRE PROCESSING BAB 3)

Adeniyi, M. O., Ekum, M. I., C, I., S, O. A., A, A. J., Oke, S. I. and B, M. M. (2020), 'Dynamic model of COVID-19 disease with exploratory data analysis', Scientific African 9, e00477. URL: <https://linkinghub.elsevier.com/retrieve/pii/S2468227620302155> (REFERENSI UNTUK EDA BAB 3)

https://openlibrary.telkomuniversity.ac.id/pustaka/files/100440/jurnal_eproc/kombinasi-algoritma-agglomerative-clustering-dan-k-means-untuk-segmentasi-pengunjung-website.pdf (REFERENSI UNTUK MODELLING BAB 3)

D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," Matrix J. Manaj. Teknol. dan Inform., vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662 (UNTUK EVALUASI MODELLING BAB 3)