

Penerapan Algoritma K-Means Clustering Pada Jumlah Fasilitas Kesehatan Menurut Pemerintah Provinsi DKI Jakarta

Aryo Satrio Wibowo¹, Indurasmi Dian Mulyastuti²
Institut Teknologi dan Bisnis Bank Rakyat Indonesia
aryosatriowibowo@gmail.com¹,
indurasmi.dianm@gmail.com²

Abstract

Fasilitas Kesehatan adalah tempat yang menyediakan pelayanan Kesehatan, seperti klinik dan rumah sakit. Kuantitas dan kualitas fasilitas ini di suatu daerah atau negara merupakan salah satu parameter yang umum digunakan mengukur nilai kemakmuran dan kualitas hidup daerah tersebut. Dalam melakukan penelitian ini, algoritma yang digunakan adalah algoritma K-Means Clustering melalui aplikasi Orange dan mengelompokkannya menjadi 3 cluster yaitu tinggi, sedang dan rendah. Data yang digunakan adalah data kabupaten berdasarkan jumlah fasilitas kesehatan di DKI Jakarta tahun 2020, tersedia 24 data dan data penelitian ini diambil dari BPS (Badan Pusat Statistik). Hasil yang diperoleh dari penelitian ini adalah cluster 1 memiliki 6 data, cluster 2 memiliki 5 data dan cluster 3 memiliki 12 data. Penelitian ini bertujuan untuk mengetahui jumlah fasilitas kesehatan di kabupaten DKI Jakarta yang tergolong tinggi, sedang dan rendah.

Keyword: *Clustering, Fasilitas Kesehatan, K-Means, Orange*

Pendahuluan

Masalah kesehatan yang ada di dalam masyarakat terutama di negara-negara berkembang seperti Indonesia dipengaruhi oleh dua faktor yaitu aspek fisik dan aspek non fisik. Aspek fisik seperti sarana kesehatan dan pengobatan penyakit, sedangkan yang kedua adalah aspek non fisik yang menyangkut masalah kesehatan seperti perilaku Manusia dalam bentuk pengetahuan, sikap dan tindakan [1].

Pelayanan kesehatan adalah hak dasar masyarakat yang wajib terpenuhi dalam rangka pembangunan sarana kesehatan. Kondisi fasilitas kesehatan di Indonesia saat ini masih perlu diperhatikan oleh pemerintah, khususnya di kabupaten DKI Jakarta. Walaupun beberapa kabupaten sudah cukup memadai, dengan pesatnya pertumbuhan masyarakat yang melebihi jumlah fasilitas akan menjadi sebuah tantangan bagi rumah sakit, puskesmas dan penyedia sarana kesehatan lainnya untuk mengimbangi permintaan sehingga banyak masyarakat merasa kesulitan untuk menjalani pola hidup yang sehat dan mengobati penyakitnya. Oleh karena itu, pemerintah diharapkan untuk dapat memperbaiki dan meningkatkan kualitas serta kuantitas fasilitas kesehatan di kabupaten DKI Jakarta. Penelitian ini dilakukan guna untuk mengklasifikasikan kabupaten yang memiliki sarana kesehatan secara cepat dan efisien.

Tinjauan Pustaka

Algoritma K-Means merupakan metode clustering yang paling sederhana dan umum. Hal ini dikarenakan K-Means mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien [2]. Algoritma K-Means merupakan algoritma klasterisasi yang mengelompokkan data berdasarkan titik pusat klaster (Centroid) terdekat dengan data. Tujuan dari K-Means adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antar klaster. Ukuran kemiripan yang digunakan dalam klaster adalah fungsi jarak. Sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data terhadap titik centroid.

- Fasilitas Kesehatan

Fasilitas kesehatan adalah suatu alat dan/atau tempat yang digunakan untuk menyelenggarakan upaya pelayanan kesehatan, baik promotif, preventif, kuratif maupun rehabilitatif yang dilakukan oleh pemerintah daerah dan/atau masyarakat. (Peraturan Pemerintah RI Nomor 47 Tahun 2016)

Fasilitas Kesehatan di Indonesia terdiri dari beberapa tingkatan yang di dalamnya terdapat beberapa jenis. Jenis fasilitas Kesehatan terbagi menjadi dua, yaitu perseorangan dan masyarakat. Jenis fasilitas Kesehatan terdiri dari tempat praktik mandiri tenaga Kesehatan, Pusat Kesehatan Masyarakat (Puskesmas), klinik, rumah sakit dan lain sebagainya. Sedangkan, tingkatan fasilitas Kesehatan terdiri dari tingkat pertama, kedua dan ketiga, berikut penjelasannya.

1. Tingkat Pertama

Fasilitas kesehatan tingkat pertama memberikan pelayanan kesehatan dasar seperti Puskesmas, klinik dan tempat praktik mandiri tenaga kesehatan.

2. Tingkat Kedua

Fasilitas kesehatan tingkat kedua memberikan pelayanan kesehatan spesialisik atau lebih lengkap dibandingkan dengan tingkat pertama. Tingkat kedua ini terdiri dari rumah sakit kelas D dan C.

3. Tingkat Ketiga

Fasilitas kesehatan tingkat ketiga memberikan pelayanan kesehatan subspecialistik atau lebih lengkap dibandingkan dengan tingkat pertama dan kedua. Pada tingkat ketiga ini, biasanya terdiri dari rumah sakit kelas A dan B.

- Data Mining

Data Mining merupakan proses pengekstrasian informasi dari sekumpulan data yang sangat besar melalui penggunaan algoritma dan teknik penarikan dalam bidang statistik, pembelajaran mesin dan sistem manajemen basis data. *Data Mining* adalah proses menganalisa data dari perspektif yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan, memperkecil biaya pengeluaran atau bahkan keduanya [3].

- Clustering

Clustering merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan record yang memiliki kemiripan dengan yang lainnya dan memiliki ketidakmiripan dengan record dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variable target dalam pengklusteran [4].

- K-Means

Algoritma *clustering* yang masuk ke dalam kelompok *Unsupervised Learning* yang digunakan untuk mengelompokan data ke dalam beberapa kelompok dengan sistem partisi. *Unsupervised Learning* merupakan algoritma data mining untuk mencari pola dari semua variabel (atribut). Variabel (atribut) yang menjadi target/label/class tidak ditentukan (tidak ada). Algoritma *K-Means* merupakan teknik *clustering* berbasis jarak yang membagi data ke dalam beberapa *cluster* dan algoritma ini hanya bekerja pada atribut angka atau numerik.

Langkah-langkah algoritma K-Means adalah sebagai berikut:

1. Tentukan nilai K atau jumlah *cluster* pada dataset.
2. Menentukan nilai pusat (*centroid*). Penentuan nilai *centroid* pada tahap awal dilakukan secara random, sedangkan pada tahap iterasi digunakan rumus seperti dibawah:

$$V_{ij} = \frac{1}{N_i} \sum_{k=1}^N X_{kj}$$

Keterangan:

V_{ij} = *Centroid* rata-rata *cluster* ke 1 untuk variabel kj

N_i = Jumlah anggota *cluster* ke- i

i, k = Indeks dari *cluster*

j = Indeks dari variabel

X_{kj} = Nilai data ke- k variabel ke- j untuk *cluster* tersebut

- Menghitung jarak antara titik *centroid* dengan titik tiap objek menggunakan Euclidean Distance, yaitu jarak garis lurus biasa antara titik dalam ruang Euclidean, dengan rumus seperti dibawah:

$$De = \sqrt{(x_i - x_s)^2 + (y_i - y_t)^2}$$

Keterangan:

De = Euclidean Distance

i = Banyaknya objek

x,y = Koordinat objek

s,t = Koordinat centroid

- Kelompok objek berdasarkan jarak ke *centroid* terdekat
- Ulangi langkah ke-2 hingga ke-4, lakukan iterasi hingga centroid bernilai optimal

- Orange

Orange data mining adalah paket perangkat pemrograman visual berbasis komponen untuk visualisasi data machine learning, data mining, dan analisis data. Pemrograman visual diimplementasikan melalui antarmuka di mana alur kerja dibuat dengan menghubungkan widget yang telah ditentukan sebelumnya atau yang dirancang pengguna.

Metode Penelitian

A. Pengumpulan Data

Dalam penelitian ini, digunakan data mengenai jumlah fasilitas kesehatan menurut kabupaten DKI Jakarta diambil dari website <https://jakarta.bps.go.id> sebanyak 6 kabupaten pada tahun 2020.

| Faskes | Jumlah Fasilitas Kesehatan Menurut Kabupaten/Kota di Provinsi DKI Jakarta | | | | | | | | | |
|---------------------|---|------|-----------------|------|---------------|------|---------------|------|---------------|------|
| | Kep. Seribu | | Jakarta Selatan | | Jakarta Timur | | Jakarta Pusat | | Jakarta Barat | |
| | 2018 | 2020 | 2018 | 2020 | 2018 | 2020 | 2018 | 2020 | 2018 | 2020 |
| Tempat Tidur | 14 | 10 | 6720 | 4745 | 6362 | 6104 | 6349 | 5662 | 5111 | 4055 |
| Puskesmas Kelurahan | 6 | 4 | 69 | 67 | 77 | 75 | 36 | 34 | 65 | 67 |
| Puskesmas Kecamatan | 2 | 2 | 10 | 10 | 10 | 10 | 8 | 8 | 8 | 8 |
| Rumah Sakit | 1 | 1 | 53 | 57 | 46 | 54 | 36 | 38 | 27 | 29 |

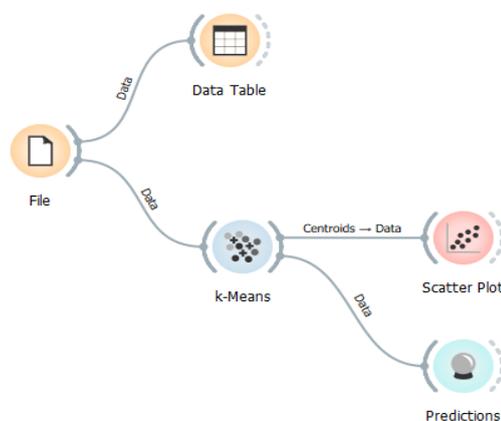
| Jumlah Fasilitas Kesehatan Menurut Kabupaten/Kota di Provinsi DKI Jakarta | | | | | | | | | | | | |
|---|-----------------|------|---------------|------|---------------|------|---------------|------|---------------|------|--------------|-------|
| Kabupaten | Jakarta Selatan | | Jakarta Timur | | Jakarta Pusat | | Jakarta Barat | | Jakarta Utara | | Jumlah Total | |
| | 2018 | 2020 | 2018 | 2020 | 2018 | 2020 | 2018 | 2020 | 2018 | 2020 | 2018 | 2020 |
| 10 | 6720 | 4745 | 6362 | 6104 | 6349 | 5662 | 5111 | 4055 | 3999 | 3091 | 28555 | 23667 |
| 4 | 69 | 67 | 77 | 75 | 36 | 34 | 65 | 67 | 43 | 41 | 296 | 288 |
| 2 | 10 | 10 | 10 | 10 | 8 | 8 | 8 | 8 | 6 | 6 | 44 | 44 |
| 1 | 53 | 57 | 46 | 54 | 36 | 38 | 27 | 29 | 27 | 30 | 190 | 209 |

Sumber: jakarta.bps.go.id

Gambar 1. Jumlah Fasilitas Kesehatan Menurut Kabupaten DKI Jakarta

B. Pengolahan Data

Pada tahap ini, digunakan software Orange untuk mengolah data dari website <https://jakarta.bps.go.id>, proses pengolahan serta workflow terdapat pada gambar berikut ini.



Gambar 2. Proses Pengolahan Data Melalui *Orange*

- Data Table

Data Table adalah representasi dalam memori dari tabel database tunggal yang memiliki kumpulan baris dan kolom. *Data Table* hanya mengambil satu *TableRow* pada satu waktu *DataSet* [5].

- Scatter Plot

Scatter Plot adalah penggunaan titik-titik untuk mewakili nilai untuk dua variable numerik yang berbeda. Posisi setiap titik pada sumbu horizontal dan vertical menunjukkan nilai untuk titik data individual. *Scatter Plot* digunakan untuk mengamati hubungan antar variable [6].

Hasil dan Pembahasan

A. Penginputan Data

Dataset yang kami gunakan pada penelitian kali ini yaitu kami peroleh dari website resmi <https://jakarta.bps.go.id> (Sumber 2020).

| Tahun | Wilayah | Fasilitas Kesehatan | Jumlah |
|-------|-----------------|---------------------|--------|
| 2020 | Kep. Seribu | Rumah Sakit | 1 |
| 2020 | Kep. Seribu | Tempat Tidur | 10 |
| 2020 | Kep. Seribu | Puskesmas Kecamatan | 2 |
| 2020 | Kep. Seribu | Puskesmas Kelurahan | 4 |
| 2020 | Jakarta Selatan | Rumah Sakit | 57 |
| 2020 | Jakarta Selatan | Tempat Tidur | 4745 |
| 2020 | Jakarta Selatan | Puskesmas Kecamatan | 10 |
| 2020 | Jakarta Selatan | Puskesmas Kelurahan | 67 |
| 2020 | Jakarta Timur | Rumah Sakit | 54 |
| 2020 | Jakarta Timur | Tempat Tidur | 6104 |
| 2020 | Jakarta Timur | Puskesmas Kecamatan | 10 |
| 2020 | Jakarta Timur | Puskesmas Kelurahan | 75 |
| 2020 | Jakarta Pusat | Rumah Sakit | 38 |
| 2020 | Jakarta Pusat | Tempat Tidur | 5662 |
| 2020 | Jakarta Pusat | Puskesmas Kecamatan | 8 |
| 2020 | Jakarta Pusat | Puskesmas Kelurahan | 34 |
| 2020 | Jakarta Barat | Rumah Sakit | 29 |
| 2020 | Jakarta Barat | Tempat Tidur | 4055 |
| 2020 | Jakarta Barat | Puskesmas Kecamatan | 8 |
| 2020 | Jakarta Barat | Puskesmas Kelurahan | 67 |
| 2020 | Jakarta Utara | Rumah Sakit | 30 |
| 2020 | Jakarta Utara | Tempat Tidur | 3091 |
| 2020 | Jakarta Utara | Puskesmas Kecamatan | 6 |
| 2020 | Jakarta Utara | Puskesmas Kelurahan | 41 |

Gambar 3. Dataset Jumlah Fasilitas Kesehatan



Gambar 4. Penginputan Data di Orange

B. Menentukan Jumlah Cluster

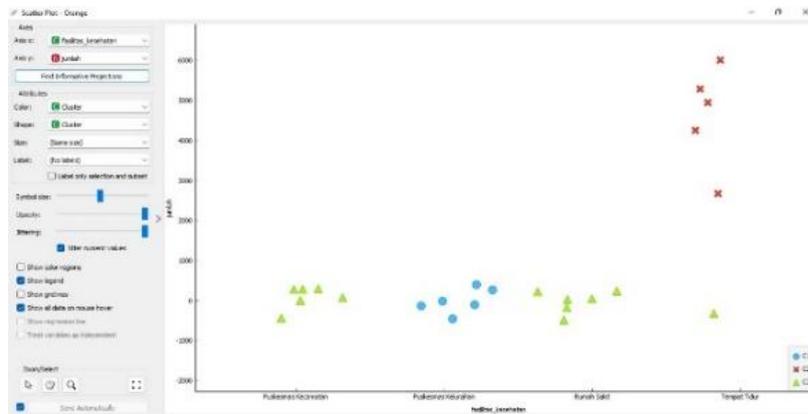
Penentuan nilai hasil cluster dilihat dari jarak terdekat antar objek data. Sebelum menghitung jarak terdekatnya, harus ditentukan terlebih dahulu jumlah cluster centroid. Pada penelitian kali ini kami menggunakan Cluster 1 (Rendah), 2 (Sedang) dan 3 (Tinggi).

| Cluster | Silhouette | tahun | wilayah | fasilitas_kesehat | jumlah |
|---------|------------|-------|-----------------|-------------------|--------|
| C1 | 0.575428 | 2020 | Kep. Seribu | Puskesmas Kel... | 4 |
| C1 | 0.58083 | 2020 | Jakarta Selatan | Puskesmas Kel... | 67 |
| C1 | 0.58083 | 2020 | Jakarta Timur | Puskesmas Kel... | 75 |
| C1 | 0.580817 | 2020 | Jakarta Pusat | Puskesmas Kel... | 34 |
| C1 | 0.58083 | 2020 | Jakarta Barat | Puskesmas Kel... | 67 |
| C1 | 0.580822 | 2020 | Jakarta Utara | Puskesmas Kel... | 41 |
| C2 | 0.645023 | 2020 | Jakarta Selatan | Tempat Tidur | 4745 |
| C2 | 0.654196 | 2020 | Jakarta Timur | Tempat Tidur | 6104 |
| C2 | 0.654307 | 2020 | Jakarta Pusat | Tempat Tidur | 5662 |
| C2 | 0.628909 | 2020 | Jakarta Barat | Tempat Tidur | 4055 |
| C2 | 0.582873 | 2020 | Jakarta Utara | Tempat Tidur | 3091 |
| C3 | 0.540632 | 2020 | Kep. Seribu | Rumah Sakit | 1 |
| C3 | 0.500017 | 2020 | Kep. Seribu | Tempat Tidur | 10 |
| C3 | 0.540635 | 2020 | Kep. Seribu | Puskesmas Kec... | 2 |
| C3 | 0.532542 | 2020 | Jakarta Selatan | Rumah Sakit | 57 |
| C3 | 0.532573 | 2020 | Jakarta Selatan | Puskesmas Kec... | 10 |
| C3 | 0.532545 | 2020 | Jakarta Timur | Rumah Sakit | 54 |
| C3 | 0.532574 | 2020 | Jakarta Timur | Puskesmas Kec... | 10 |
| C3 | 0.532556 | 2020 | Jakarta Pusat | Rumah Sakit | 38 |
| C3 | 0.532573 | 2020 | Jakarta Pusat | Puskesmas Kec... | 8 |
| C3 | 0.532562 | 2020 | Jakarta Barat | Rumah Sakit | 29 |
| C3 | 0.532575 | 2020 | Jakarta Barat | Puskesmas Kec... | 8 |
| C3 | 0.532561 | 2020 | Jakarta Utara | Rumah Sakit | 30 |
| C3 | 0.532575 | 2020 | Jakarta Utara | Puskesmas Kec... | 6 |

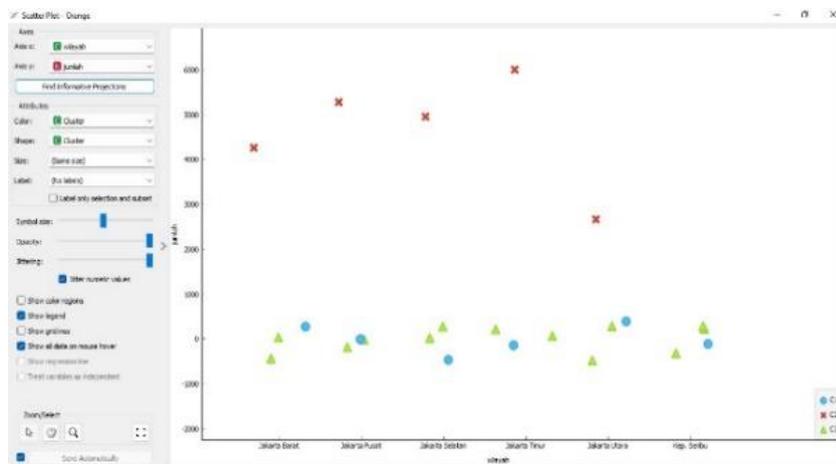
Gambar 5. Contoh Hasil Cluster 1, 2, dan 3

C. Melakukan *Plotting*

Setelah kita menentukan jumlah cluster, kita akan melakukan metode *Plotting*. *Plotting* yang dilakukan adalah membandingkan wilayah dengan jumlah dan juga fasilitas_kesehatan dengan jumlah pada masing-masing *cluster*.



Gambar 6. *Plotting* fasilitas_kesehatan dengan Jumlah



Gambar 7. *Plotting* Wilayah dengan Jumlah

Kesimpulan

Berdasarkan penelitian yang kami lakukan, kita akan mendapatkan hasil bahwa fasilitas kesehatan yang ada di Pemerintahan Provinsi DKI Jakarta menghasilkan 3 cluster, dengan fasilitas kesehatan paling sedikit berada di Kepulauan Seribu dengan jumlah 1 Rumah Sakit, 4 Puskesmas Kelurahan, 2 Puskesmas Kecamatan, dan 10 Tempat tidur. Dari hasil berikut dapat disimpulkan bahwa masih kurangnya fasilitas kesehatan di Kepulauan Seribu, hal ini dapat dijadikan gambaran bagi Pemerintah di Pemerintahan Provinsi DKI Jakarta agar menambah fasilitas kesehatan di Kepulauan Seribu terlebih daerah tersebut daerah terpencil yang jauh dari daratan pusat kota yang mana harus lebih diperhatikan agar masyarakat disana tidak perlu jauh-jauh menyebrang laut untuk berobat di pusat kota.

Daftar Pustaka

[1] Mayangsarim, Sekar D.S., Solikhun & Irawan. “Pengelompokkan Jumlah Desa/Kelurahan yang Memiliki Sarana Kesehatan menurut Provinsi dengan Menggunakan Metode K-Means Cluster”. Konferensi Nasional Teknologi Informasi dan Komputer (KOMIK). Volume III, pp.370-377, 2019.

[2] Yulia Darnita, Rozali Toyib, Yayang K. “Penerapan Metode K-Means Clustering Pada Aplikasi Android Pada Tanaman Obat Herbal”. Volume VII Nomor 2, 2020.

- [3] Muchammad Farrell Putro, Edhi Prayitno, Juarni Siregar dan Muhammad Muharrom. “Penerapan Data Mining Dengan Naïve Bayes Untuk Klasifikasi Siswa Sekolah Menengah Atas Dalam Penentuan Perguruan Tinggi”. Volume VI Nomor 2, 2021.
- [4] Indra Setiawan. “Penerapan Metode Clustering Untuk Memetakan Potensi Tanaman Kedelai di Jawa Tengah dengan Algoritma Fuzzy C-Means”. Teknik Informatika Universitas Dian Nuswantoro Semarang.
- [5] gelbvicassociationnebraska.org, 24 Juni 2022, <https://gelbvicassociationnebraska.org/what-is-a-datatable>.
- [6] dotedu.id, 24 Juni 2022, <https://dotedu.id/apa-itu-scatter-plot/>.
- [7] jakarta.bps.go.id, 18 Juni 2022, <https://jakarta.bps.go.id>.