



Pengelompokan Provinsi Di Indonesia Berdasarkan Kabupaten/Kota Yang Memiliki Sarana Kesehatan Menggunakan Algoritma *Partitioning Around Medoid* Dengan Metode *Davies Bouldin Index*

Fikril Mubarak*, Hardian Oktavianto, Qurrota A'yun

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Email: fikrilmubarak00@gmail.com*, hardian@unmuhjember.ac.id, qurrota.ayun@unmuhjember.ac.id

ABSTRAK

Sarana kesehatan merupakan sarana utama dalam pemenuhan kebutuhan masyarakat akan kesehatan di setiap daerah, untuk itu sarana kesehatan harus terletak pada posisi yang strategis dan tersebar merata diseluruh daerah. Jumlah penduduk di Indonesia yang tersebar di 34 provinsi yang mencapai 267,7 juta jiwa berdasarkan badan pusat statistik nasional pada tahun 2018 yang tentunya membutuhkan pelayanan kesehatan yang memadai. Pelayanan kesehatan dapat dilihat dari ketersediaan sarana pelayanan yang tersedia di masing - masing wilayah. Permasalahan tersebut membuat pemerataan sarana kesehatan belum maksimal sehingga masih menjadi kendala besar bagi dunia kesehatan di Indonesia. Pada penelitian ini membahas tentang pengelompokan provinsi di Indonesia berdasarkan kabupaten/kota yang memiliki sarana kesehatan menggunakan algoritma *Partitioning Around Medoids* (PAM) dengan metode *Davies Bouldin Index*, diperoleh cluster terbaik dengan hasil 2 cluster berdasarkan validasi *Davies Bouldin Index* sebesar 0,223 dengan skenario 2 cluster sampai 10 cluster. Sedangkan jumlah anggota pada masing-masing cluster yaitu cluster 1 terdapat 31 provinsi dan cluster 2 terdapat 3 provinsi.

Kata Kunci: Sarana Kesehatan, Clustering, *Partitioning Around Medoids* (PAM), *Davies Bouldin Index*

ABSTRACT

*Health facilities are the main means of meeting the needs of the community for health in each area, therefore health facilities must be located in a strategic position and are evenly distributed throughout the region. The total population in Indonesia which is spread across 34 provinces currently reaches 267.7 million people based on the national statistics center in 2018 which of course requires adequate health services, health services can be seen from the availability of service facilities available in each region. These problems have made the distribution of health facilities not maximized so that it is still a big obstacle for the world of health in Indonesia. This study discusses the grouping of provinces in Indonesia based on districts / cities that have health facilities using the *Partitioning Around Medoids* (PAM) algorithm with the *Davies Bouldin Index* method, the best cluster is obtained with 2 clusters based on the *Davies Bouldin Index* validation of 0.223 with a scenario of 2 clusters to 10 clusters. Meanwhile, the number of members in each cluster is 31 provinces in cluster 1 and 3 provinces in cluster 2.*

Keywords: *Health facilities, Clustering, Partitioning Around Medoids Davies Bouldin Index*

1. PENDAHULUAN

Sarana kesehatan adalah tempat yang digunakan untuk menyelenggarakan upaya kesehatan (Aprella, 2016). Undang-undang Nomor 36 tahun 2009 tentang kesehatan menyatakan bahwa fasilitas pelayanan kesehatan merupakan tempat yang digunakan sebagai penyelenggaraan kesehatan baik yang bersifat promotif, preventif, kuratif maupun rehabilitatif yang dilakukan oleh pemerintah daerah maupun pusat dan juga masyarakat (UU RI No 36, 2009). Berdasarkan pada undang-undang tersebut maka sudah jelas bahwa pentingnya pelayanan kesehatan bagi masyarakat memiliki sifat yang mutlak. Jumlah penduduk di Indonesia yang tersebar di 34 provinsi saat ini yang mencapai 267,7 juta jiwa berdasarkan badan pusat statistik nasional pada tahun 2018 yang tentunya membutuhkan pelayanan kesehatan yang memadai. Pelayanan kesehatan dapat dilihat dari ketersediaan sarana pelayanan yang tersedia di masing-masing wilayah. Permasalahan tersebut membuat pemerataan sarana kesehatan belum maksimal sehingga masih menjadi kendala besar bagi dunia kesehatan di Indonesia.

Pada penelitian sebelumnya yang telah dilakukan oleh Putri (2018) dengan judul “Implementasi Algoritma *Partitioning Around medoids* untuk mengelompokkan SMA/MA se-kota Pekanbaru”. Pada penelitian tersebut untuk mengetahui hasil dari penerapan algoritma *Partitioning Around Medoids* dan mengevaluasi hasil pengelompokan SMA/MA se-Kota Pekanbaru dengan menghitung kualitas cluster

menggunakan silhouette index. Pengujian data dilakukan dengan menghasilkan kelompok data yang dimiliki cluster 1 lebih tinggi dari pada cluster dengan dibuktikannya pada cluster 2 masih banyak sekolah yang nilai per setiap indikator soal dibawah batas standar minimal kelulusan Ujian Nasional. Hal ini menjadi bahan evaluasi bagi sekolah yang berada di cluster 2 untuk dapat meningkatkan nilai daya serap hingga nilai ujian nasional meningkat.

Kelebihan K-medoids yaitu menggunakan objek sebagai perwakilan (medoid) pusat cluster untuk tiap cluster. Algoritma K-medoids dapat membantu dalam pengelompokan provinsi mana saja yang masih kurang dalam pemerataan sarana kesehatan. Berdasarkan kelebihan tersebut, maka penulis ingin melakukan penelitian terhadap sarana kesehatan di Indonesia di tahun 2018 dengan judul "Pengelompokan provinsi di Indonesia berdasarkan kabupaten/kota yang memiliki sarana kesehatan menggunakan Algoritma Partitioning Around medoids (PAM) dengan metode daivies bouldin index (DBI)". Pada penelitian ini terdapat 8 atribut yang penulis gunakan, yaitu rumah sakit, rumah sakit bersalin, poliklinik, puskesmas, puskesmas pembantu, apotek, jumlah penduduk (ribu) dan luas wilayah (km²). Selain menggunakan K-medoids, penulis juga melakukan kombinasi dengan menggunakan metode davies bouldin index. Metode ini adalah metode yang digunakan untuk mengukur validitas cluster pada suatu metode clustering.

2. KAJIAN PUSTAKA

A. Sarana Kesehatan

Undang-undang Nomor 36 tahun 2009 tentang kesehatan menyatakan bahwa fasilitas pelayanan kesehatan merupakan tempat yang digunakan sebagai penyelenggaraan kesehatan baik yang bersifat promotif, preventif, kuaratif maupun rehabilitatif yang dilakukan oleh pemerintah daerah maupun pusat dan juga masyarakat (UU RI No 36, 2009).

Kesehatan sebagai salah satu unsur kesejahteraan umum harus diwujudkan sesuai dengan cita-cita bangsa Indonesia sebagaimana dimaksud dalam pembukaan UUD 1945 melalui pembangunan nasional yang berkesinambungan.

Berdasarkan Pancasila dan UUD 1945, dari beberapa pengertian diatas dapat diambil kesimpulan bahwa sarana kesehatan merupakan sarana utama dalam pemenuhan kebutuhan masyarakat akan kesehatan di setiap daerah untuk itu sarana kesehatan harus terletak pada posisi yang strategis dan tersebar merata diseluruh daerah.

B. Data Mining

Menurut Han dan Kamber (2011), data mining adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar. Sedangkan menurut Linoff dan Berry (2011). Data mining adalah suatu pencarian dan analisa dari jumlah data yang sangat besar dan bertujuan untuk mencari arti dari pola dan aturan. Dari beberapa teori yang dijabarkan oleh para ahli, bahwa data mining adalah suatu pencarian dan analisa pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik dengan tujuan mengekstrak informasi dan knowledge yang akurat dan berpotensi, serta dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting.

C. Clustering

Larose (2015) clustering merupakan suatu proses pengelompokan record, observasi, atau mengelompokkan kelas yang memiliki kesamaan objek. Perbedaan clustering dengan klasifikasi yaitu tidak adanya variabel target dalam melakukan pengelompokan pada proses clustering. Clustering sering dilakukan sebagai awal dalam proses data mining.

Prasetyo (2012) clustering dapat melakukan pemisahan atau segmentasi pada data ke dalam yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip satu sama lain. Lebih besar kemiripannya dalam kelompok dan lebih besar perbedaannya diantara kelompok yang lain. Dari

beberapa pengertian diatas bahwa clustering merupakan pengelompokan (cluster) data yang besar yang memiliki kesamaan objek dalam kelompok data.

D. *Partitioning Around Medoids*

Menurut Santoso, Februariyanti, dan Hery, (2016) Algoritma Partitioning Around Medoids (PAM) adalah algoritma pengelompokan yang berkaitan dengan algoritma k-means. Dengan kata lain kedua algoritma ini memecah data set menjadi kelompok-kelompok dan kedua algoritma ini berusaha untuk meminimalkan kesalahan tetapi algoritma Partitioning Around Medoids (PAM) bekerja dengan menggunakan medoids yang merupakan entitas dari dataset yang mewakili kelompok dimana data dimasukan.

Algoritma K-Medoids merupakan teknik partisi klasik dari clustering yang melakukan klasterisasi dataset objek n ke dalam k cluster yang dikenal sebagai a priori (Abhishek & Purnima, 2013). Algoritma ini beroperasi pada prinsip untuk minimalkan jumlah kesamaan antara setiap objek dan titik referensi yang sesuai. Algoritma K-Medoids dapat dilakukan dengan langkah-langkah sebagai berikut (Bhat, 2014):

- 1) Inialisasi pusat cluster sebanyak k (jumlah cluster).
- 2) Hitung setiap objek ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance. Perhitungan Euclidian Distance menggunakan persamaan sejumlah kelompok (cluster) menurut karakteristik tertentu dalam pengelompokan label dari setiap data belum diketahui dan dengan pengelompokan diharapkan dapat diketahui kelompok data untuk kemudian diberi label. Tujuannya adalah objek-objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip satu sama lain. Lebih besar kemiripannya dalam kelompok dan lebih besar perbedaannya diantara kelompok yang lain. Dari beberapa pengertian diatas bahwa clustering merupakan pengelompokan (cluster) data yang besar yang memiliki kesamaan objek dalam kelompok data.

$$\text{Total cost} = \sum \sqrt{\sum n}$$

Dengan:

n = jumlah sebuah data

k = indeks data

x_k = nilai atribut ke-k dari x

y_k = nilai atribut ke-k dari y

- 3) Setelah menghitung jarak Euclidian Distance, inialisasikan pusat cluster baru secara acak pada masing-masing objek sebagai kandidat non-medoids.
- 4) Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat non-medoids.
- 5) Hitung total simpangan (S) dengan menghitung total distance baru – total distance lama. Jika S < 0 maka tukar objek dengan data cluster non-medoids untuk membentuk sekumpulan k objek baru sebagai medoids.

$$S = \text{Total cost baru} - \text{Total cost lama} \quad (1)$$

Dengan:

S = selisih

C_i = centroid cluster ke-i

d(x_j c_j) = jarak setiap data terhadap centroid

- 6) Menghitung Sum of square between cluster (SSB), yaitu persamaan yang digunakan untuk mengetahui separasi antar cluster yang dihitung menggunakan persamaan sebagai berikut:

$$SSB_{i,j} = d(c_i, c_j) \quad (2)$$

- 7) Menghitung jumlah ratio (rasio). Setelah nilai SSW dan SSB diperoleh kemudian melakukan pengukuran rasio (R_{i,j}) untuk mengetahui nilai perbandingan antara cluster ke-i dan cluster ke-j nilai rasio dihitung menggunakan persamaan Davies Bouldin Index. Davies Bouldin Index (DBI) adalah metode yang digunakan untuk mengukur validitas cluster pada suatu metode clustering.

Dalam penelitian ini DBI digunakan untuk melakukan validasi data pada setiap cluster. Langkah-langkah untuk menghitung nilai Davies Bouldin Index (DBI).

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (3)$$

Dengan:

SSW_i = Sum of Square within-i (matrikkoehesi dalam sebuah cluster ke i)

SSW_j = Sum of Square within-j (matrikkoehesi dalam sebuah cluster ke j)

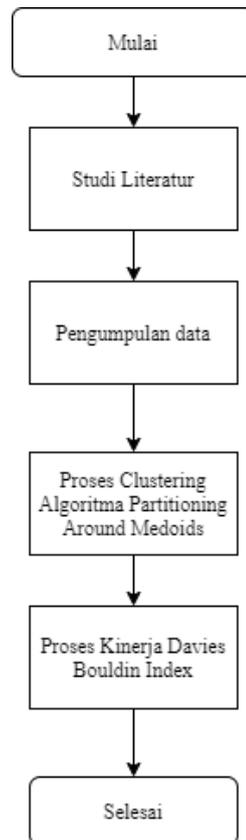
SSB_{ij} = Sum of square Between - ij (separasi antar cluster ke-i dan ke-j)

Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai Davies Bouldin Index (DBI).

$$(xk - yk)^2 \quad (4)$$

3. METODE PENELITIAN

Penelitian ini menggunakan algoritma Partitioning Around Medoids (PAM) clustering dengan teknik performance-nya menggunakan Davies Bouldin Index dalam penentuan cluster terbaik, memiliki tahapan sebagai berikut:



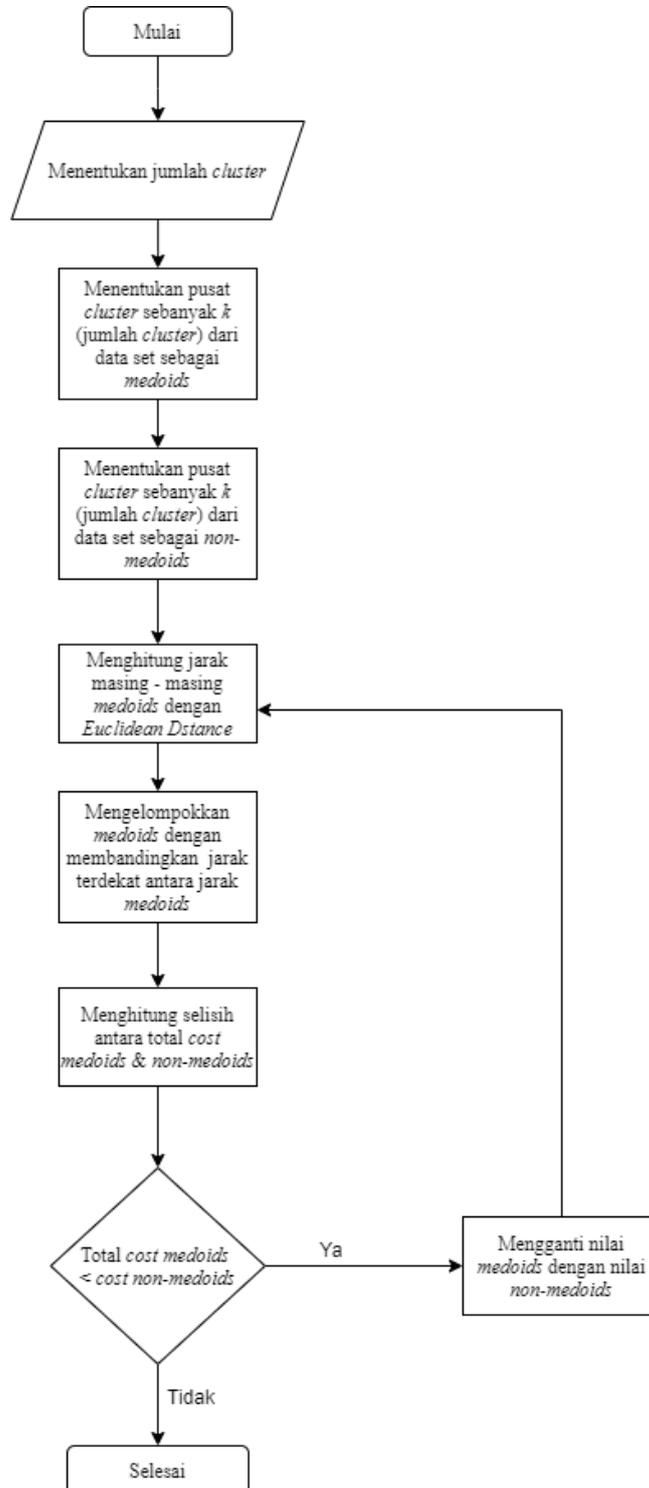
Gambar 1. Diagram Alur Penelitian

A. Studi Literatur

Tahapan pertama dari penelitian ini untuk mencari dan mempelajari masalah yang akan diteliti kemudian menentukan ruang lingkup masalah, latar belakang dan mempelajari beberapa literatur yang berkaitan dengan permasalahan dan bagaimana mencari solusi dari masalah tersebut. Untuk mencapai tujuan yang ditentukan maka penulis perlu mempelajari beberapa literatur yang digunakan kemudian literatur tersebut diseleksi untuk ditentukan sebagai literatur yang akan digunakan dalam penelitian.

B. Metode Analisis Data

Tahapan awal penelitian ini metode analisa data yang digunakan adalah Partitioning Around Medoids (PAM) merupakan teknik cluster atau mengelompokkan dari beberapa objek yang mewakili (medoids) di dalam pusat cluster untuk setiap cluster. Kelebihan dari PAM yaitu untuk mengatasi kelemahan dari algoritma K-means yang sensitif terhadap nois atau outlier dan objek dengan nilai yang besar yang memungkinkan menyimpang dari distribusi data.



Gambar 2. Flowchart Algoritma PAM

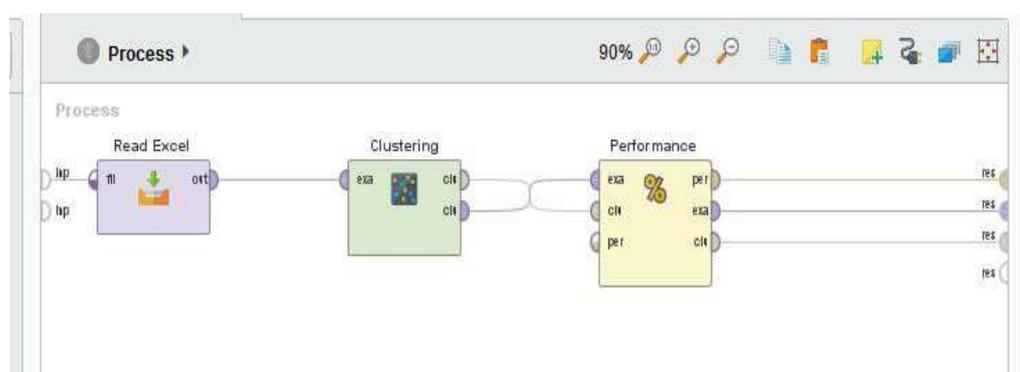
C. Pengumpulan Data

Pengumpulan data dilakukan berdasarkan studi literatur dari situs resmi Badan Pusat Statistik Indonesia yakni <https://www.bps.go.id> untuk melengkapi data yang diperlukan. Data yang digunakan untuk penelitian ini adalah data provinsi di Indonesia berdasarkan kabupaten/kota yang memiliki sarana kesehatan pada tahun 2018 sebanyak 272 data.

4. HASIL DAN PEMBAHASAN

Hasil yang akan diperoleh dari pengujian data, data tersebut akan diolah menggunakan algoritma Partitioning Around Medoids (PAM). Berikut hasil dan pembahasan dari Algoritma Partitioning Around Medoids (PAM) untuk mengelompokkan provinsi di Indonesia berdasarkan kabupaten/kota yang memiliki sarana kesehatan berdasarkan 8 atribut yaitu rumah sakit, rumah sakit bersalin, poliklinik, puskesmas, puskesmas pembantu, apotek, jumlah penduduk (ribu) dan luas wilayah (km²). Data diperoleh dari Badan Pusat Statistik (BPS) di Indonesia pada tahun 2018.

A. RapidMiner Studio



Gambar 3. Proses Kinerja Partitioning Around Medoids Pada RapidMiner

Berdasarkan gambar 3 terdapat beberapa operator yang digunakan. Berikut fungsi dari operator-operator tersebut.

- 1) Read Exel: Operator ini digunakan untuk memuat data dari sheet pada Microsoft Exel
- 2) Clustering: Operator ini melakukan pengelompokan menggunakan metode clustering k- medoids. Pada penelitian ini jumlah cluster yang akan digunakan sebagai pengujian yaitu 2 sampai 10 cluster.
- 3) Performance: Operator performance yang digunakan yaitu cluster distance performance dimana operator ini digunakan untuk evaluasi kinerja metode k-medoids berdasarkan nilai davies bouldin index.

B. Penentuan Jumlah Cluster Terbaik Setelah melakukan Proses Cluster

Setelah melalui proses cluster dengan menggunakan algoritma PAM, kemudian dilakukan proses dengan menggunakan metode DBI untuk penentuan cluster terbaik. Berikut adalah hasil dari metode DBI.

Tabel 1. Hasil Nilai Metode DBI

Cluster	Nilai DBI	Anggota Cluster
2	0,223	Cluster 1: 31 dan Cluster 2: 3
3	0,456	Cluster 1: 24 Cluster 2: 7 dan Cluster 3: 3
4	0,809	Cluster 1: 11

Cluster	Nilai DBI	Anggota Cluster
		Cluster 2: 7 Cluster 3: 13 dan Cluster 4: 3
5	0,660	Cluster 1: 6 Cluster 2: 5 Cluster 3: 13 Cluster 4: 7 dan Cluster 5: 3
6	0,969	Cluster 1: 8 Cluster 2: 6 Cluster 3: 5 Cluster 4: 5 Cluster 5: 7 dan Cluster 6: 3
7	0,666	Cluster 1: 5 Cluster 2: 5 Cluster 3: 15 Cluster 4: 3 Cluster 5: 3 Cluster 6: 3 dan Cluster 7: 2
8	0,602	Cluster 1: 5 Cluster 2: 7 Cluster 3: 3 Cluster 4: 3 Cluster 5: 6 Cluster 6: 3 Cluster 7: 5 Cluster 8: 2
9	3,206	Cluster 1: 3 Cluster 2: 8 Cluster 3: 7 Cluster 4: 7 Cluster 5: 2 Cluster 6: 2 Cluster 7: 1 Cluster 8: 1 dan Cluster 9: 3
10	0,925	Cluster 1: 4 Cluster 2: 5 Cluster 3: 6 Cluster 4: 7 Cluster 5: 3 Cluster 6: 2 Cluster 7: 4 Cluster 8: 1 Cluster 9: 1 dan Cluster 10: 1

Hasil dari validasi cluster adalah cluster yang memiliki nilai terendah. Pada tabel 4 ditunjukkan bahwa nilai DBI terdapat di cluster 2 yang memiliki nilai DBI 0,223. Jadi, cluster 2 merupakan cluster terbaik.

Dari hasil karakteristik pengelompokan sarana kesehatan didapatkan hasil fitur pada cluster 2 yaitu rumah sakit, rumah sakit bersalin, poliklinik, puskesmas, puskesmas pembantu, apotek, jumlah penduduk (ribu) dan luas wilayah (km²) yang memiliki jumlah bangunan sarana kesehatan yang tinggi dibandingkan dengan cluster 1 yang memiliki jumlah bangunan sarana kesehatan yang rendah.

5. KESIMPULAN

Hasil dari penggunaan algoritma Partitioning Around Medoids dalam pengelompokan provinsi di Indonesia berdasarkan kabupaten/kota yang memiliki sarana kesehatan diperoleh cluster terbaik adalah cluster 2 yang memiliki nilai terendah berdasarkan validasi Davies Bouldin Index yaitu 0,223 dengan skenario 2 cluster sampai 10 cluster. Hasil dari pengelompokan dari 2 cluster pada cluster 1 terdapat 31 provinsi terdiri dari Aceh, Sumatra Utara, Sumatra Barat, Riau, Jambi, Sumatra Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat dan Papua. Pada cluster 2 terdapat 3 provinsi terdiri dari Jawa Barat, Jawa Tengah dan Jawa Timur. Hasil cluster profiling menunjukkan tingkat pemerataan sarana kesehatan di setiap provinsi di Indonesia pada cluster 1 memiliki karakteristik data dengan anggota-anggota data yang jumlah sarana kesehatannya rendah dan pada cluster 2 memiliki karakteristik data dengan anggota-anggota data yang jumlah sarana kesehatannya tinggi yang dipengaruhi oleh jumlah penduduk di setiap provinsi di Indonesia.

Bagi penelitian selanjutnya disarankan untuk menggunakan bahasa pemrograman lain dalam pengembangan sistem, seperti bahasa pemrograman Java pada sistem Android, atau bahasa pemrograman lainnya. Dalam proses validasi cluster dapat dikembangkan untuk mencari cluster terbaik dengan menggunakan alternatif lain seperti silhouette coefficient, elbow, dan gap statistic. Algoritma Partitioning Around Medoids pada penelitian ini dapat dikembangkan dengan data yang lebih baru pada studi kasus yang berbeda.

6. DAFTAR PUSTAKA

- Abhishek, P. & Purnima, S. (2013). New Approach for K-means and K-Medoids Algorithm. International journal of Computer Application Technology and Research. Vol. 2. India.
- Aprella, Q. A.P. 2016. Pengaruh Pola Sebaran Sarana dan Prasarana Kesehatan Terhadap Aksesibilitas Pelayanan Kesehatan Masyarakat di Kabupaten Tegal.
- Bhat, A. (2014). *K-Medoids Clustering Using Partitioning Around Medoids for Performing Face Recognition*. International Journal of Soft Computing, Mathematics and Control (IJSCMC), Vol. 3, No.3.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Fransisco: Morgan Kauffmann Publisher.
- Larose, Daniel, T. and Larose, Chantal D. 2015. Data mining and Predictive Analytics. Second Edition, John Wiley & Sons. Edition, John Wiley & Sons.
- Linoff, G. S., & Berry, M. J. (2011). *Data mining Techniques for Marketing, Sales, Customer Relationship Management*. UnitedStates of America: Wiley Publishing, Inc.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. Andi. Yogyakarta.
- PUTRI, A. (2018). Implementasi Algoritma Partitioning Around Medoids Untuk Pengelompokan Sma/Ma Se-Kota Pekanbaru (Dissertation, Riau: Universitas Islam Negeri Sultan Syarif Kasim Riau).
- Putri, A. Q. A., Puji, H., & Mochammad, A. (2018). Pengaruh Pola Sebaran Sarana dan Prasarana Kesehatan Terhadap Aksesibilitas Pelayanan Kesehatan Masyarakat di Kabupaten Tegal Tahun 2016. *Geo-Image Journal*, 7(1), 31-38.
- Santoso, D.B., Februriyanti, & Henry. 2016. "Algoritma Partitioning Around Medoids (PAM) Clustering untuk Melihat Gambaran Umum Skripsi Mahasiswa". *Jurnal Teknologi Informasi DINAMIK* Vol. 21, No. 1, Januari 2016. Universitas Stikubank, Semarang.
- Undang-Undang Republik Indonesia Nomor 36 Tahun 2009. Tentang Kesehatan. Jakarta.