

PENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN KABUPATEN/KOTA YANG MEMILIKI SARANA KESEHATAN MENGGUNAKAN ALGORITMA *PARTITIONING AROUND MEDOID* DENGAN METODE *DAVIES BOULDIN INDEX*

Fikril Mubarak¹, Hardian Oktavianto², Qurrota A'yun³

Program Studi Teknik Informatika, Fakultas Teknik Universitas Muhammadiyah Jember

Jln. Karimata No. 49 Jember Kode Pos 68121

E-mail: fikrilmubarak00@gmail.com

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 pengelompokkan 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*. Dan 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*. And the number of members in each cluster, namely cluster 1 there are 31 provinces and cluster 2 there are 3 provinces.

Keywords: Health facilities, *clustering*, *Partitioning Around Medoids* (PAM), *Davies Bouldin Index*.

1. PENDAHULUAN

Sarana kesehatan adalah tempat yang digunakan untuk menyelenggarakan upaya kesehatan (Aprilla, Q. A.P. 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 di lakukan 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 pengelompokkan 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* 2 dengan dibuktikan nya 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 pengelompokkan 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 "Pengelompokkan provinsi di Indonesia berdasarkan kabupaten/kota yang memiliki sarana kesehatan menggunakan Algoritma *Partitioning Around medoids* (PAM) dengan metode *davies 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. TINJAUAN PUSTAKA

2.1 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, kuratif maupun rehabilitatif yang di lakukan 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 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

2.2 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.

2.3 Clustering

Iarose (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 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 di antara kelompok yang lain. Dari beberapa pengertian diatas bahwa *clustering* merupakan pengelompokan (*cluster*) data yang besar yang memiliki kesamaan objek dalam kelompok data.

2.4 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 dimasukkan.

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 meminimalkan jumlah kesamaan antara setiap objek dan titik referensi yang sesuai. Algoritma *K-Medoids* dapat dilakukan dengan langkah-langkah sebagai berikut (Bhat, 2014):

- a. Inisialisasi pusat *cluster* sebanyak *k* (jumlah *cluster*).
- b. Hitung setiap objek ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance*. Perhitungan *Euclidian Distance* menggunakan persamaan

$$\text{Total cost} = \sum \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \dots \dots \dots (2.1)$$

Dengan:

n = jumlah sebuah data

k = indeks data

x_k = nilai atribut ke-*k* dari *x*

y_k =nilai atribut ke-k dari y

- c. Setelah menghitung jarak *Euclidian Distance*, inialisasikan pusat *cluster* baru secara acak pada masing – masing objek sebagai kandidat *non-medoids*.
- d. Hitung jarak setiap objek yang berada pada masing – masing *cluster* dengan kandidat *non- medoids*.
- e. 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*.

$$S = \text{Total cost baru} - \text{Total cost lama} \dots\dots(2.2)$$

Dengan:

S = selisih

Total *cost* baru = Jumlah *cost non-medoids*

Total *cost* lama = Jumlah *cost medoids*

- f. Ulangi langkah c – e hingga tidak terjadi perubahan pada *medoid*, sehingga di dapatkan *cluster* beserta anggota *cluster* masing – masing

2.5 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*. Adapun langkah-langkah untuk menghitung *Davies Bouldin Index* (Sujacka, 2019) yaitu:

1. Menghitung *Sum of Square within cluster* (SSW) *Sum of Square within cluster* merupakan persamaan yang digunakan untuk mengetahui matrik kohesi dalam sebuah *cluster* ke *i* yang dirumuskan sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j; c_i) \dots\dots\dots(2.3)$$

Dengan :

m_i =jumlah data dalam *cluster* ke-*i*

C_i = centroid *cluster* ke-*i*

$d(x_j; c_j)$ = jarak setiap data terhadap centroid

2. Menghitung *Sum of square between cluster* (SSB) *Sum of square Between cluster* (SSB) adalah persamaan yang digunakan untuk mengetahui separasi antar *cluster* yang dihitung menggunakan persamaan sebagai berikut:

$$SSB_{i,j} = d(c_i, c_j) \dots\dots\dots(2.4)$$

3. 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 sebagai berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \dots\dots\dots(2.5)$$

Dengan :

SSW_i = *Sum of Square within-i* (matrik kohesi dalam sebuah *cluster* ke *i*)

SSW_j = *Sum of Square within-j* (matrik kohesi dalam sebuah *cluster* ke *j*)

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

4. Menghitung nilai *Davies Bouldin Index* (DBI) Nilai rasio yang diperoleh tersebut digunakan untuk mencari nilai *davies bouldin index* (DBI) dengan menggunakan persamaan berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \dots\dots(2.6)$$

Dengan :

k = jumlah *cluster* yang ditentukan.

R_{ij} = rasio *cluster* ke-*i* dan *cluster* ke-*j*

Skema *clustering* yang optimal menurut indeks pengukuran *Davies Bouldin Index* adalah *cluster* yang memiliki nilai indeks terkecil atau minimal (Salazar,dkk.2012).

2.6 Rapid Miner

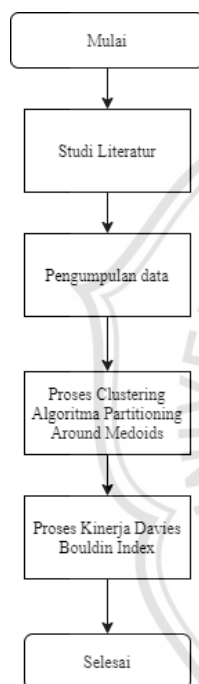
Rapid miner menyediakan GUI (*Graphic User Interface*) untuk merancang sebuah *pipeline* analitis

.GUI akan menghasilkan file XML (*Extensible Markup language*) yang didefinisikan proses analitis keinginan pengguna untuk ditetapkan ke data. File ini kemudian dibaca oleh rapid miner untuk menjalankan analisis secara otomatis.

3. METODOLOGI

3.1 Tahapan 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.



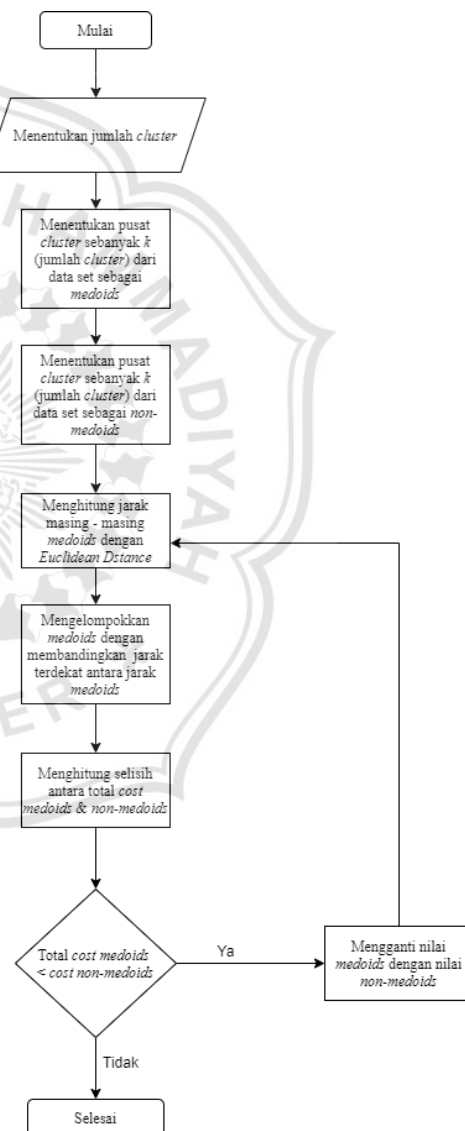
Gambar 1. Diagram Alur Penelitian

3.2 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.

3.3 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

3.4 Pengumpulan Data

Pengumpulan data dilakukan berdasarkan studi literatur dari situs resmi Badan Pusat Statistik

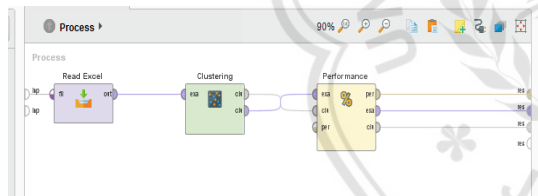
Indonesia yakni <https://www.bps.go.id> untuk melengkapi data yang diperlukan. data yang digunakan untuk penelitian ini adalah data provinsi di Indonesia berdasarkan kabupaten/kota yang memiliki sarana kesehatan pada tahun 2018 sebanyak 272 data.

4. HASIL DAN PEMBAHASAN

4.1 Data Pengujian

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.

4.2 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.

- Read Excel*: Operator ini digunakan untuk memuat data dari sheet pada Microsoft Exel
- Clustering* : Operator ini melakukan pengelompokkan menggunakan metode *clustering k-medoids*. Pada penelitian ini jumlah *cluster* yang akan digunakan sebagai pengujian yaitu 2 sampai 10 *cluster*.

- Performance* : Operator *performance* yang digunakan yaitu *cluster distance performance* dimana operator ini digunakan untuk evaluasi kinerja metode *k-medoids* berdasarkan nilai *davies bouldin index*.

4.3 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 4 hasil Nilai Metode DBI

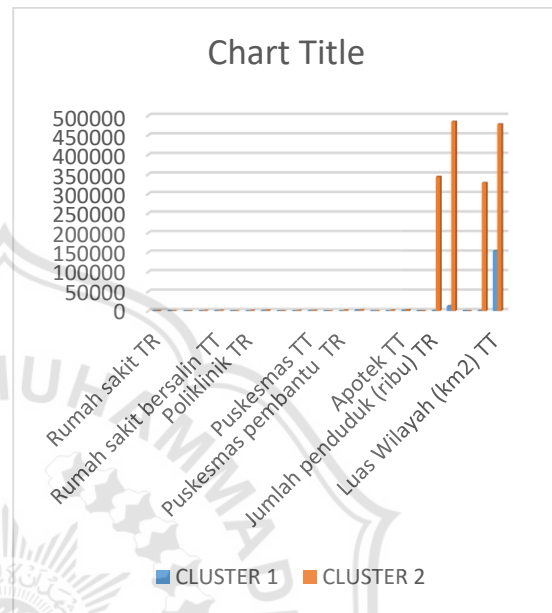
Cluster	Nilai DBI	Anggota Cluster
2	0,223	Cluster 1: 31 dan Cluster 2 : 3
3	0,456	Cluster 1: 24, Cluster 2: 7 dan Cluster 3 : 3
4	0,809	Cluster 1: 11, Cluster 2: 7, Cluster 3: 13 dan Cluster 4 : 3
5	0,660	Cluster 1: 6, Cluster 2: 5, Cluster 3: 13, Cluster 4: 7 dan Cluster 5 : 3
6	0,969	Cluster 1: 8, 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 dan 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

Pada metode DBI, nilai *cluster* yang diambil untuk digunakan sebagai *cluster* terbaik

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.

4.4 Hasil Cluster Profiling



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 DAN SARAN

5.1 Kesimpulan

1. 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*.

2. 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, Kepulauan Bangka Belitung, Kepulauan 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.

5.2 Saran

1. Dalam proses validasi *cluster* dapat dikembangkan untuk mencari *cluster* terbaik dengan menggunakan alternatif lain seperti *silhouette coefficient*, *elbow*, *gap statistic*, dll.
2. Algoritma *Partitioning Around Medoids* pada penelitian ini dapat dikembangkan dengan data yang lebih baru pada studi kasus yang berbeda.

6 DAFTAR PUSTAKA

- [1] Aprella, Q. A.P. 2016. *Pengaruh Pola Sebaran Sarana dan Prasarana Kesehatan Terhadap Aksesibilitas Pelayanan Kesehatan Masyarakat di Kabupaten Tegal*.
- [2] Aeni, N.R. 2020. *Algoritma Partitioning*

Around Medoids dalam mengelompokkan provinsi di Indonesia berdasarkan index kinerja davies bouldin pada kasus penyakit HIV. Universitas Muhammadiyah Jember.

- [3] Christie, A.D., Baskoro, D.A., Ambarwati, I., Wicaksana, I.W.S. 2013 “*Belajar Data Mining Dengan Rapid Miner*”. Jakarta: Gramedia Pustaka.
- [4] Frebruriyanti, H., Santoso, D.B., 2016. *Algoritma Partitioning Around Medoids (PAM) Clustering untuk Melihat Gambaran Umum Skripsi Mahasiswa*. Universitas Stikubank
- [5] Hermawati, F.A. 2013. *Data Mining*. Surabaya: Andi Offset.
- [6] Helma, S.S., 2019. *Clustering pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K-Means*. Universitas Islam Negeri Sultan Syarif Kasim Riau
- [7] Hasugian, P. 2018. “*Penerapan data mining Untuk Klasifikasi Produk Menggunakan Algoritma K-Means*”. Jurnal Mantik Penusa. Volume 2, No.2. Teknik Informatika STMIK Pelita Nusantara, Medan.
- [8] Hardiyanti, F. 2019. *Penerapan Data Mining menggunakan algoritma K-Medoids untuk mengelompokkan penanganan kasus diare di Indonesia*. STIKOM Tunas Bangsa Pematangsiantar.
- [9] Hartanto, D., & Hansun, S. (2014). *Implementasi data mining dengan Algoritma C4.5 untuk memprediksi tingkat kelulusan mahasiswa*. Jurnal UTMATICS, 1,15-20.
- [10] Kementerian Kesehatan Republik Indonesia, 2018. *Profil Kesehatan Indonesia*. Jakarta.
- [11] Kurnawan, I., Marisa F., Purnomo, D. 2018. *Implementasi data mining dengan Algoritma Apriori untuk memprediksi tingkat kelulusan mahasiswa*. Universitas Widyagama Malang.
- [12] Larose, Daniel, T. and Larose, Chantal D. 2015. *data mining and Predictive Analytics*. Second

Edition, John Wiley & Sons.

[13] Putri, A. 2018. *Implementasi Algoritma Partitioning Around medoids untuk mengelompokkan SMA/MA se-kota Pekanbaru*.

[14] 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.

[15] Silitonga, D.A., Windarto, A.P., Hartama, D., Sumarno. 2019. "Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi". STIKOM Tunas Bangsa, Pematangsiantar Indonesia.

[16] Sujacka, R. 2019. *Peningkatan Akurasi Algoritma K-Means dengan Clustering Purity sebagai Titik Pusat Cluster Awal (Centroid)*. [pdf] repository.usu.ac.id.

[17] Undang-Undang Republik Indonesia Nomor 36 Tahun 2009. *Tentang Kesehatan*. Jakarta.

[18] Wirdasari, D. & A.Calam,. *Penerapan Data Mining Untuk Mengolah Data Penempatan Buku Di Perpustakaan Smk Ti Pab 7 lubuk Pakam Dengan Metode Association Rule*. *Jurnal SAINTIKOM*, 10 (2), 150, 2011.

