ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 15, Nomor 2, Tahun 2025, Halaman 312 - 319
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI JAWA TENGAH MENGGUNAKAN ALGORITMA PAMDAN ALGORITMA DBSCAN BERDASARKAN INDIKATOR KELUARGA SEHAT

Inez Clarissa Nababan[1*], **Puspita Kartikasari**[1], **Ardiana Alifatus Sa'adah**[1]
[1] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***email**: inezclarissanababan11@gmail.com

DOI: 10.14710/j.gauss.15.2.312-319

Article Info:
Received: 2024-09-17
Accepted: 2025-12-30
Available Online: 2025-08-31

Keywords:
Healthy Family; Cluster Analysis; PAM; DBSCAN; Silhouette Index.

Abstract: The healthy family index is an index used to provide an overview of the health conditions in a family. The health status of families in Central Java Province varies in each district/city, so health development priorities are also different. Based on indications of a healthy family, this study seeks to classify Central Java Province's districts and cities to determine the high/low quality of healthy families in each district/city in order to aid the government in maximizing its efforts to enhance health. The combination of district/city in Central Java Province was carried out using cluster analysis using the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm and the Partitioning Around Medoids (PAM) algorithm. The outcomes of grouping using the PAM algorithm obtained 4 clusters where cluster one had 16 cluster members, cluster two had 10 cluster members, cluster three had 8 cluster members, and cluster four had 1 cluster members. Grouping using the DBSCAN algorithm produces 2 clusters and 3 noise. There are many members in each cluster, namely cluster one has 29 cluster members and cluster two has 3 cluster members. Comparing the grouping results using the silhouette index value, it was found that the DBSCAN algorithm was the best algorithm with a value of the silhouette index of 0.1425. The description of the characteristics obtained is that cluster two is a cluster that has a lower average of 6 healthy family indicators compared to cluster one.

1. PENDAHULUAN

Kesehatan merupakan unsur pokok pembangunan dalam mencapai kesejahteraan masyarakat (Dinkes Jateng, 2019). Berdasarkan hal tersebut, pembangunan kesehatan sangat penting untuk dilakukan. Pembangunan kesehatan merupakan suatu usaha yang dilakukan agar kesehatan masyarakat dapat meningkat. Kementerian Kesehatan mencanangkan suatu program yang berupaya dapat mengembangkan agar setiap individu memiliki perilaku hidup sehat, lingkungan sehat, dan pemahaman tentang kesehatan dengan baik.

Menurut Hair *et al*. (2010), analisis klaster dapat diartikan sebagai suatu metode yang bertujuan untuk mempartisi data berdasarkan karakteristik data tersebut. Algoritma PAM merupakan suatu algoritma klastering yang menggunakan objek *medoids* sebagai pusat klaster (Kaur, *et al*., 2014). Algoritma PAM dapat melakukan pengelompokan terhadap data yang menyimpang (Han dan Kamber, 2006). Data yang menyimpang dapat disebut juga dengan *outlier*. Algoritma DBSCAN merupakan suatu algoritma klastering yang mengelompokkan objek berdasarkan kepadatan objek. Menurut Mumtaz dan Duraiswamy (2010), algoritma DBSCAN dapat membentuk klaster dari data yang mengandung *noise* dan *outlier*.

Validasi klaster perlu dilakukan dalam analisis klaster agar diperoleh klaster terbaik. Penelitian ini membahas tentang pengelompokan kabupaten/kota di Provinsi Jawa Tengah menggunakan algoritma PAM dan algoritma DBSCAN berdasarkan indikator keluarga sehat dengan validasi klaster yang digunakan adalah *silhouette index*.

Hasil penelitian yang dilakukan diharapkan dapat menggambarkan kabupaten/kota Provinsi Jawa Tengah berdasarkan indikator keluarga sehat, sehingga dengan adanya informasi tersebut dapat diketahui tinggi/rendahnya kualitas keluarga sehat pada daerah tersebut guna membantu pemerintah dalam menentukan prioritas program mana yang harus mendapat dukungan dan dalam menangani permasalahan kesehatan untuk mengoptimalkan upaya pembangunan kesehatan.

2. TINJAUAN PUSTAKA

Indeks keluarga sehat bertujuan untuk mengevaluasi kesehatan keluarga secara keseluruhan. Kementerian Kesehatan membuat suatu program untuk mengatasi masalah kesehatan dan memperluas akses keluarga terhadap perawatan medis. Program ini dilakukan dengan pendekatan keluarga (Kemenkes RI, 2017).

Data yang menyimpang dari pola umum model disebut juga dengan pencilan (*outlier*) (Sembiring, 1995). Adanya pencilan dapat menyebabkan hasil analisis tidak akurat, sehingga kondisi populasi tidak tercerminkan secara akurat (Hair, *et al*., 2010). Pada data multivariat, untuk mendeteksi data pencilan bisa menggunakan perhitungan jarak kuadrat Mahalanobis. Jarak kuadrat Mahalanobis dirumuskan sebagai berikut:

$$d_{MD}^2(h) = (\boldsymbol{x_h} - \bar{\boldsymbol{x}})^T \Sigma^{-1} (\boldsymbol{x_h} - \bar{\boldsymbol{x}}),\ h = 1, 2, \ldots, n \quad (1)$$

dengan: $\boldsymbol{x_h}$ = vektor objek ke-h untuk setiap variabel, $\bar{\boldsymbol{x}}$ = vektor rata-rata dari setiap variabel, Σ^{-1} = matriks kovarian. Menurut Johnson dan Wichern (2007), suatu data teridentifikasi sebagai data pencilan, apabila data tersebut mempunyai nilai jarak kuadrat Mahalanobis $\left(d_{MD}^2(h)\right)$ > nilai distribusi *chi-kuadrat* $(\chi_{1-\alpha,p}^2)$.

Variabel penelitian harus distandarisasi apabila memungkinkan untuk menghindari masalah yang dihasilkan dari penggunaan satuan yang berbeda. Tujuan standarisasi data adalah untuk menyamakan satuan, sehingga nilai standar tidak lagi bergantung pada satuan pengukuran melainkan menjadi nilai baku. Standarisasi yang paling umum adalah konversi setiap variabel terhadap nilai standar (*Z-score*). Nilai standar (*Z-score*) dapat dirumuskan sebagai berikut:

$$Z_{hj} = \frac{x_{hj} - \bar{x}_j}{s_j};\ h = 1, 2, \ldots, n;\ j = 1, 2, \cdots, p \quad (2)$$

dengan: x_{hj} = objek ke-h pada variabel ke-j, $\bar{x}_j$ = rata-rata variabel ke-j, s_j =standar deviasi variabel ke-j

Menurut Hair *et al.* (2010), analisis klaster digunakan untuk mempartisi objek ke dalam suatu kelompok berdasarkan karakteristik objek tersebut. Asumsi yang harus terpenuhi dalam analisis klaster yaitu asumsi non-multikolinearitas. Ketika dua variabel atau lebih memiliki hubungan linier sempurna, maka hal ini dapat disebut sebagai multikolinearitas. (Gujarati, 1978). Uji asumsi non-multikolinearitas bisa menggunakan nilai *Variance Inflation Factor* (VIF). Nilai VIF dirumuskan sebagai berikut:

$$VIF_j = \frac{1}{\left(1 - R_j^2\right)} \quad (3)$$

dengan:

R_j^2 = koefisien determinasi variabel ke-*j*. Suatu data dapat dikatakan terjadi multikolinearitas, apabila mempunyai nilai $VIF_j \geq 10$.

Algoritma PAM merupakan suatu algoritma yang bertujuan untuk mengelompokan sekumpulan data menjadi beberapa klaster. Menurut Kaufman dan Rousseeuw (1990), klaster akan terbentuk melalui perhitungan kedekatan jarak *medoids* dengan objek pengamatan. Langkah-langkah algoritma PAM:

1. Masukkan data $(\boldsymbol{x_h})$, berupa matriks $\boldsymbol{X}$ dengan ukuran *n* x *p*
2. Tetapkan *k* sebagai banyaknya klaster yang ingin dibentuk
3. Pilih *medoids* awal $\left(\boldsymbol{m_i^{c,t}}\right)$ secara acak sebagai pusat klaster sebanyak *k* dari *n* objek pada iterasi *t*, dengan *c* = 1,2, ⋯, *k*
4. Hitung jarak antara objek dan *medoids* awal pada iterasi *t*
$$d_{euc}(x_h, m_i^{c,t}) = \sqrt{\sum_{j=1}^{p}(x_{h,j} - m_{i,j}^{c,t})^2}$$
5. Tentukan anggota klaster dari setiap objek (s_h^t) berdasarkan jarak terdekat objek dengan *medoids* dari iterasi *t*
$$s_h^t = min\{d_{euc}(x_h, m_i^{c,t})\}$$
6. Hitung total jarak terdekat (S_t) pada iterasi *t*
$$S_t = \sum_{h=1}^{n} s_h^t$$
7. Pilih *medoids* baru $\left(\boldsymbol{m_i^{c,t}}\right)$ secara acak sebagai pusat klaster sebanyak *k* dari *n* objek pada iterasi *t*+1, dengan *c* = 1,2, ⋯, *k*, melakukan kembali langkah (4) sampai langkah (6)
8. Hitung selisih antara total jarak terdekat objek pada iterasi *t*+1 dengan iterasi *t* $\left(S_{total\ jarak}\right)$
$$S_{total\ jarak} = S_{(t+1)} - S_t$$

Apabila nilai $S_{total\ jarak} > 0$, maka iterasi akan berhenti dan apabila nilai $S_{total\ jarak} < 0$, maka *medoids* baru tersebut menjadi *medoids* awal untuk iterasi selanjutnya serta iterasi dilanjutkan dengan mengulang langkah (7) yaitu memilih *medoids* baru sampai langkah (8) yaitu menghitung selisih total jarak, hingga didapatkan nilai $S_{total\ jarak} > 0$.

Algoritma DBSCAN merupakan algoritma klastering yang mengelompokkan objek berdasarkan kepadatan objek. Konsep kepadatan dalam algoritma DBSCAN diartikan sebagai jumlah objek yang berada dalam radius Eps. Menurut Prasetyo (2012), algoritma DBSCAN menghasilkan tiga macam status, yaitu inti, batas, dan *noise*. Menurut Purwanto (2012), algoritma DBSCAN memiliki dua parameter yang harus ditentukan secara *trial and error*, yaitu MinPts dan Eps. Penentuan nilai parameter tersebut dapat dilakukan dengan membuat grafik *k-dist*. Pada grafik *k-dist*, titik yang mengalami perubahan tajam akan menjadi nilai Eps dan *k* menjadi nilai MinPts. Langkah-langkah algoritma DBSCAN:

1. Masukkan data $(\boldsymbol{x_h})$, berupa matriks $\boldsymbol{X}$ dengan ukuran *n* x *p*
2. Tetapkan nilai *k* untuk membuat grafik *k-dist*
3. Tandai semua objek sebagai "*unvisited*"
4. Tentukan satu objek *i* sebagai titik pusat untuk iterasi *t* $(\boldsymbol{c_i^t})$ secara acak dari semua objek, lalu memberi tanda "*visited*"
5. Pada iterasi *t*, hitung jarak masing-masing objek terhadap titik pusat

$$d_{euc}(x_h, c_i^t) = \sqrt{\sum_{j=1}^{p}(x_{h,j} - c_{i,j}^t)^2}$$

6. Ambil semua objek yang *density-reachable* terhadap titik pusat berdasarkan hasil perhitungan jarak pada iterasi *t*, lalu memberi tanda "*visited*", dimana objek yang *density-reachable* terhadap titik pusat didapat dari sekumpulan titik *directly density-reachable*, yaitu
$$d_{euc}(x_h, c_i^t) \leq Eps$$
7. Uji apakah jumlah titik yang diambil tersebut berada dalam *Neighborhood* dari titik pusat $(Neps(i))$, dimana
$$|Neps(i)| \geq MinPts$$
Jika iya, maka titik-titik tersebut dapat dikatakan sebagai hasil iterasi *t*. Jika tidak, maka titik-titik tersebut dapat dikatakan sebagai *noise*
8. Tentukan satu objek *i* sebagai titik pusat pada iterasi selanjutnya dengan syarat yaitu memilih titik yang terjauh dalam Eps-*Neighborhood* dari titik pusat $(Neps(i))$ pada iterasi sebelumnya dan titik yang bukan termasuk titik batas pada titik pusat iterasi sebelumnya.
9. Jika objek *i* adalah titik batas dan objek yang *density-reachable* terhadap titik pusat tidak ada, maka tentukan titik pusat selanjutnya ke objek yang masih bertanda "*unvisited*"
10. Ulangi langkah (4) sampai (9) hingga semua objek sudah bertanda "*visited*"

Hasil pengelompokan yang didapatkan dari algoritma klastering dapat dievaluasi dengan melakukan validasi klaster. Tujuan validasi klaster dilakukan adalah untuk memperoleh hasil pengelompokan yang terbaik. Terdapat dua kriteria validasi klaster yang dapat dilakukan, yaitu kriteria internal dan kriteria eksternal. Penelitian ini akan menggunakan validasi kriteria internal, yaitu *silhouette index*. Menurut Rousseeuw (1987), validasi *silhouette index* menghitung rata-rata untuk setiap objek. Nilai *silhouette index* untuk semua objek dirumuskan sebagai berikut:

$$SI = \frac{1}{n}\sum_{h=1}^{n} SI_h^c, \text{ dengan: } SI_h^c = \frac{b_h^c - a_h^c}{max\{a_h^c; b_h^c\}} \quad (4)$$

dengan: *SI* = *silhouette index* global, SI_h^c = *silhouette index* objek ke-*h* pada klaster *c*, a_h^c = nilai rata-rata jarak objek ke-*h* dengan semua objek dalam klaster *c*, b_h^c = nilai minimum rata-rata jarak objek ke-*h* dengan semua objek selain klaster *c*.

3. METODE PENELITIAN

Data penelitian yang digunakan merupakan data indikator keluarga sehat periode Januari sampai Desember tahun 2021. Data diperoleh dari Badan Pusat Statistik Provinsi Jawa Tengah dan Dinas Kesehatan Provinsi Jawa Tengah. Variabel penelitian yang digunakan yaitu persentase peserta KB aktif, persentase ibu melahirkan di fasilitas kesehatan, persentase imunisasi dasar lengkap pada bayi, persentase balita yang telah menerima imunisasi dasar lengkap, persentase cakupan kepesertaan JKN, persentase bayi yang menerima ASI eksklusif, persentase angka keberhasilan pengobatan TBC, persentase rumah tangga yang memiliki fasilitas sendiri untuk buang air besar, persentase rumah tangga yang memiliki akses terhadap sumber air minum layak konsumsi, serta persentase desa yang melaksanakan STBM. Tahapan-tahapan pengolahan data yang dilakukan yaitu:

1. Deteksi data pencilan menggunakan perhitungan jarak kuadrat Mahalanobis.

2. Melakukan standarisasi data.
3. Hitung nilai VIF untuk uji asumsi non-multikolinearitas.
4. Analisis klaster dengan algoritma PAM dan algoritma DBSCAN.
4. Uji validasi klaster dengan *silhouette index* pada klaster yang telah dibentuk menggunakan algoritma PAM dan algoritma DBSCAN.
5. Bandingkan hasil pengelompokan antara algoritma PAM dan algoritma DBSCAN.
6. Interpretasi karakteristik berdasarkan klaster yang terbentuk.

4. HASIL DAN PEMBAHASAN

Data pencilan dapat dideteksi menggunakan perhitungan jarak kuadrat Mahalanobis. Suatu data teridentifikasi sebagai data pencilan, apabila data tersebut mempunyai nilai jarak kuadrat Mahalanobis $\left(d_{MD}^2(h)\right)$ > nilai distribusi *chi-kuadrat* $\left(\chi_{1-\alpha,p}^2\right)$. Nilai distribusi *chi-kuadrat* dengan $\alpha = 5\%$ diperoleh sebesar 18,3070. Berdasarkan perhitungan jarak kuadrat Mahalanobis diketahui terdapat 3 daerah teridentifikasi sebagai data pencilan, yaitu Kabupaten Banjarnegara, Kabupaten Demak, dan Kabupaten Pemalang.

Standarisasi data dilakukan untuk menyeragamkan satuan data penelitian. Standarisasi data yang digunakan adalah standarisasi *z-score*. Asumsi yang perlu terpenuhi dalam analisis klaster adalah asumsi non-multikolinearitas. Uji asumsi non multikolinearitas dilakukan dengan menghitung nilai VIF. Tabel 1 berikut menunjukkan nilai VIF yang diperoleh.

Tabel 1. Nilai VIF

Variabel	VIF
X_1	1,1447
X_2	1,7920
X_3	1,2470
X_4	1,2669
X_5	1,6408
X_6	1,3503
X_7	1,5101
X_8	1,7018
X_9	1,9444
X_{10}	1,1172

Berdasarkan nilai VIF yang diperoleh dapat diketahui bahwa variabel X_1 sampai X_{10} mempunyai nilai VIF < 10, maka dapat disimpulkan data memenuhi asumsi non-multikolinearitas.

Pengelompokan menggunakan algoritma PAM dengan k = 4 dapat diketahui bahwa banyak anggota masing-masing klasternya, yaitu pada objek pengamatan ke-28 sebagai pusat klaster satu memiliki 16 anggota klaster, objek pengamatan ke-35 sebagai pusat klaster dua memiliki 10 anggota klaster, objek pengamatan ke-3 sebagai pusat klaster tiga memiliki 8 anggota klaster, serta objek pengamatan ke-21 sebagai pusat klaster empat memiliki 1 anggota klaster.

Pengelompokan menggunakan algoritma DBSCAN dengan parameter MinPts = 2 dan Eps = 3,0 menghasilkan 2 klaster dan 3 *noise*. Banyak anggota masing-masing klasternya, yaitu pada klaster 1 memiliki 29 anggota klaster dan klaster 2 memiliki 3 anggota klaster.

Hasil pengelompokan terbaik dilihat dari nilai validasi *silhouette index* yang maksimum, sehingga dapat disimpulkan bahwa pengelompokan menggunakan algoritma DBSCAN lebih baik daripada algoritma PAM, karena pada algoritma DBSCAN diperoleh nilai *silhouette index* yang lebih besar yaitu sebesar 0,1425 dibanding nilai *silhouette index* pada algoritma PAM sebesar 0,0920. Tabel 2 berikut menunjukkan hasil pengelompokan terbaik yang diperoleh.

Tabel 2. Hasil Pengelompokan Terbaik

Klaster	Jumlah Anggota	Kabupaten/Kota
0	3	Kabupaten Banjarnegara, Kabupaten Demak, dan Kabupaten Pemalang
1	29	Kabupaten Cilacap, Kabupaten Banyumas, Kabupaten Purbalingga, Kabupaten Kebumen, Kabupaten Magelang, Kabupaten Boyolali, Kabupaten Klaten, Kabupaten Sukoharjo, Kabupaten Wonogiri, Kabupaten Karanganyar, Kabupaten Sragen, Kabupaten Grobogan, Kabupaten Blora, Kabupaten Rembang, Kabupaten Pati, Kabupaten Kudus, Kabupaten Jepara, Kabupaten Semarang, Kabupaten Kendal, Kabupaten Batang, Kabupaten Pekalongan, Kabupaten Tegal, Kabupaten Brebes, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan Kota Tegal
2	3	Kabupaten Purworejo, Kabupaten Wonosobo dan Kabupaten Temanggung

Keterangan: Klaster 0 merupakan *noise*

Rata-rata setiap klaster atau yang dikenal dengan *centroid* diperlukan untuk mengetahui karakteristik pada klaster. Tabel 3 berikut menunjukkan nilai *centroid* yang diperoleh.

Tabel 3. Nilai *Centroid*

Variabel	Klaster 1	Klaster 2
Persentase peserta KB aktif (X_1)	70,0	74,0
Persentase ibu melahirkan di fasilitas kesehatan (X_2)	99,9	99,8
Persentase imunisasi dasar lengkap pada bayi (X_3)	88,7	81,5
Persentase bayi yang menerima ASI eksklusif (X_4)	69,9	85,7
Persentase balita yang telah menerima imunisasi dasar lengkap (X_5)	69,5	76,7
Persentase angka keberhasilan pengobatan TBC (X_6)	84,4	75,6
Persentase cakupan kepesertaan JKN (X_7)	85,2	79,1
Persentase rumah tangga yang memiliki akses terhadap sumber air minum layak konsumsi (X_8)	95,0	91,2
Persentase rumah tangga yang memiliki fasilitas sendiri untuk buang air besar(X_9)	86,7	67,0
Persentase desa yang melaksanakan STBM (X_{10})	100,0	100,0

Informasi yang diperoleh berdasarkan hasil pengelompokan terbaik dan nilai *centroid*, yaitu:

a. Klaster 1

Anggota pada klaster satu yaitu Kabupaten Cilacap, Kabupaten Banyumas, Kabupaten Purbalingga, Kabupaten Kebumen, Kabupaten Magelang, Kabupaten Boyolali, Kabupaten Klaten, Kabupaten Sukoharjo, Kabupaten Wonogiri, Kabupaten Karanganyar, Kabupaten Sragen, Kabupaten Grobogan, Kabupaten Blora, Kabupaten Rembang, Kabupaten Pati, Kabupaten Kudus, Kabupaten Jepara, Kabupaten Semarang, Kabupaten Kendal, Kabupaten Batang, Kabupaten Pekalongan, Kabupaten Tegal, Kabupaten Brebes, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan

Kota Tegal. Indikator keluarga sehat yang menonjol dengan nilai rata-rata yang tinggi pada klaster satu yaitu persentase ibu melahirkan di fasilitas kesehatan, persentase imunisasi dasar lengkap pada bayi, persentase angka keberhasilan pengobatan TBC, persentase cakupan kepesertaan JKN, persentase rumah tangga yang memiliki akses terhadap sumber air minum layak konsumsi, persentase rumah tangga yang memiliki fasilitas sendiri untuk buang air besar, dan persentase desa yang melaksanakan STBM.

b. Klaster 2

Anggota pada klaster dua yaitu Kabupaten Purworejo, Kabupaten Wonosobo dan Kabupaten Temanggung. Indikator keluarga sehat yang menonjol dengan nilai rata-rata yang tinggi pada klaster dua yaitu persentase peserta KB aktif, persentase bayi yang menerima ASI eksklusif, persentase balita yang telah menerima imunisasi dasar lengkap, dan persentase desa yang melaksanakan STBM.

c. *Noise*

Hasil pengelompokan terbaik menunjukkan terdapat 3 daerah sebagai *noise* yaitu Kabupaten Banjarnegara, Kabupaten Demak, dan Kabupaten Pemalang. Hal ini berarti bahwa daerah Kabupaten Banjarnegara, Kabupaten Demak, dan Kabupaten Pemalang Surakarta merupakan daerah yang kurang padat.

5. KESIMPULAN

Pengelompokan menggunakan algoritma PAM dapat diketahui bahwa banyak anggota masing-masing klasternya, yaitu pada klaster satu memiliki 16 anggota klaster, klaster dua memiliki 10 anggota klaster, klaster tiga memiliki 8 anggota klaster, serta klaster empat memiliki 1 anggota klaster. Pengelompokan menggunakan algoritma DBSCAN menghasilkan 2 klaster dan 3 *noise*. Banyak anggota masing-masing klasternya, yaitu pada klaster satu memiliki 29 anggota klaster dan klaster dua memiliki 3 anggota klaster.

Perbandingan hasil pengelompokan menggunakan nilai *silhouette index* antara algoritma DBSCAN dan algoritma PAM, didapat bahwa algoritma DBSCAN adalah algoritma terbaik dalam melakukan pengelompokan data indikator keluarga sehat di Provinsi Jawa Tengah dengan nilai validasi *silhouette index* yaitu 0,1425.

Gambaran karakteristik berdasarkan klaster yang terbentuk yaitu klaster dua merupakan klaster yang memiliki rata-rata 6 indikator keluarga sehat yang lebih rendah jika dibandingkan dengan klaster satu, yaitu persentase ibu melahirkan di fasilitas kesehatan, persentase imunisasi dasar lengkap pada bayi, persentase angka keberhasilan pengobatan TBC, persentase cakupan kepesertaan JKN, persentase rumah tangga yang memiliki akses terhadap sumber air minum layak konsumsi, dan persentase rumah tangga yang memiliki fasilitas sendiri untuk buang air besar.

DAFTAR PUSTAKA

[DINKES JATENG] Dinas Kesehatan Provinsi Jawa Tengah. 2019. *Rencana Strategis Dinas Kesehatan Provinsi Jawa Tengah tahun 2018–2023*. Jawa Tengah: Dinas Kesehatan Provinsi Jawa Tengah.

Gujarati, D. 1978. *Ekonometrika Dasar*. Jakarta: Erlangga.

Hair, J. F., Anderson, R. E., Thatham, R. L., dan Black, W.C. 2010. *Multivariate Data Analysis 7th Edition*. New Jersey: Pearson Education Inc.

Han, J. dan Kamber, M. 2006. *Data Mining: Concepts and Techniques 2th Edition*. San Francisco: Elsevier Inc.

Johnson, R. A. dan Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis 6th*. United State of America: Pearson Education Inc.

Kaufman, L. dan Rousseeuw, P. J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis (Wiley Series in Probability and Statistics).* New York: Wiley.

Kaur, N. K., Kaur, U., dan Singh, D. 2014. K-Medoids Clustering Algorithm. *International Journal of Computer Application and Technology* (IJCAT). Vol. 1, No. 1.

[KEMENKES RI] Kementerian Kesehatan Republik Indonesia. 2017. Tersedia: https://www.kemkes.go.id/article/view/17070700004/program-indonesia-sehat-dengan-pendekatan-keluarga.html (diakses pada tanggal 01 Maret 2023).

Mumtaz, K., dan Duraiswamy, K. 2010. An Analysis on Density Based Clustering of Multi Dimensional Spatial Data. *Indian Journal of Computer Science and Engineering.* Vol. 1, No. 1, Hal: 8-12.

Prasetyo, E. 2012. *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: Andi.

Purwanto, U. Y. 2012. Spatial Hotspots Clustering of Forest and Land Fires using DBSCAN and ST-DBSCAN. *Scientific Repository (IPB)*.

Rousseeuw, P. J. 1987. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics*. Vol. 20, Hal: 53–65.

Sembiring, R. K. 1995. *Analisis Regresi*. Bandung: Penerbit ITB.