

# **Perbandingan Metode Clustering dalam Analisis Kesehatan di Indonesia tahun 2021: *K-Means, Fuzzy C-Means, dan Gaussian Mixture Model (GMM)***

Angga Fathan Rofiqy<sup>1‡</sup>, Natasha Muti Hafiza<sup>2</sup>, Arfiah Kania Sektiaruni<sup>3</sup>, Naswa Nabila Zahrani<sup>4</sup>, Kamilah Nurul Azizah<sup>5</sup>, Angel Martha Pradina P<sup>6</sup>

## **1. Pendahuluan**

Kesehatan merupakan salah satu aspek yang penting dalam pembangunan manusia, peran kesehatan ini sebagai aset utama pembangunan bangsa. *United Nations* (UN) memasukkan aspek kesehatan menjadi salah satu tujuan pada *Sustainable Development Goals* (SDGs) ketiga yang berbunyi, “Menjamin kehidupan yang sehat dan meningkatkan kesejahteraan seluruh penduduk semua usia” (BPS 2016). Menurut *World Health Organization* (WHO), kesehatan adalah keadaan sejahtera yang meliputi fisik, mental, dan sosial yang tidak hanya bebas dari sakit atau kecacatan (Jacob dan Sandjaya 2018). Kesehatan dapat direpresentasikan melalui usia panjang dan hidup yang sehat, dimana Umur Harapan Hidup (UHH) di Indonesia tahun 2021 mengalami peningkatan. Pada tahun 2021, angka harapan hidup di Indonesia adalah 71,57 tahun, yang dapat dideskripsikan sebagai anak yang baru lahir pada tahun 2021 diharapkan dapat hidup hingga usia tersebut (BPS 2022).

Kementerian Kesehatan telah menetapkan bahwa terdapat tujuh indeks pembangunan kesehatan masyarakat diantaranya pelayanan kesehatan dan kesehatan lingkungan (Utami *et al.* 2019). Fasilitas pelayanan kesehatan dapat mencakup jaminan kesehatan, puskesmas, posyandu, dan tenaga medis. Sedangkan indikator kesehatan lingkungan dapat ditinjau dengan melihat bagaimana akses ke air bersih dan sanitasi lingkungan. Namun, pada faktanya Kementerian Kesehatan pada tahun 2018 melaporkan bahwa, jumlah penyebaran puskesmas cenderung tidak merata dan hanya tersebar di kota besar yang memiliki akses yang mudah. Selain itu menurut Kementerian Kesehatan Republik Indonesia (2021) berdasarkan data sistem informasi SDM kesehatan persebaran daerah dengan puskesmas yang telah memenuhi standar tenaga jenis kesehatan yang memadai hanya sekitar 39,6%.

Dalam pemenuhan indikator kesehatan lingkungan, diketahui bahwa akses masyarakat Indonesia terhadap air bersih hanya mencapai 72,55% jumlah ini sangat kurang dari yang seharusnya 100% (Hargono *at al.* 2022). Krisis ini terjadi karena

kelemahan dalam pengelolaan air. Pengelolaan air bersih akan berkaitan pula dengan pengelolaan sanitasi, hal ini karena air dan sanitasi adalah dua hal yang tidak dapat dipisahkan saat ada air bersih maka pasti akan ada air limbah (Suryani 2020). Sanitasi merupakan salah satu metode pengelolaan air dan lingkungan dengan membangun sarana prasarana yang bertujuan meningkatkan derajat kesehatan masyarakat (Saskia et al.2020). Karena adanya ketimpangan indeks pembangunan kesehatan ini, setiap daerah akan memiliki derajat kesehatan dan mendapatkan prioritas penanganan yang berbeda. Maka dari itu, pengelompokan daerah-daerah tersebut dapat dilakukan dengan analisis statistika yaitu *clustering*.

*Clustering* merupakan metode untuk mengelompokkan data yang memiliki karakteristik sama ke dalam satu klaster. Dalam *Clustering* terdapat banyak metode, beberapa diantaranya *K-Means*, *Gaussian Mixture Model* (GMM), dan *Fuzzy C-Means*. Metode *K-Means* adalah metode *clustering* yang umum diterapkan dan sederhana dengan kuantitas besar dalam waktu penghitungan yang cepat dan efisien (Angelie, 2017). Algoritma GMM adalah salah satu jenis *soft clustering* dimana satu data *point* bisa berada pada dua atau lebih *cluster* dan cocok digunakan untuk dataset dengan masalah yang cukup kompleks (Riyono et al. 2022). Metode GMM termasuk efisien untuk pengolahan objek dalam jumlah besar (Amaluddin et al.2017).

Penerapan metode *Fuzzy C-means* menunjukkan bahwa setiap data memiliki derajat keanggotaan yang mewakili setiap *cluster*. Algoritma *Fuzzy C-means* memiliki perhitungan dan alur yang sederhana (Nugraha et al. 2020). Penelitian yang dilakukan Patel et al. 2020 melakukan perbandingan *clustering cloud workloads* dengan menggunakan dua metode yaitu *K-Means* dan GMM. Dari penelitian tersebut menghasilkan *Gaussian Mixture Model* memberikan cluster yang lebih baik, dan lebih rinci daripada *K-Means*, namun dibutuhkan waktu yang lebih lama dibanding proses *clustering* pada *K-Means*.

Menurut Ambarsari et al. 2023, hasil pengelompokan 34 data provinsi dengan indikator kesejahteraan rakyat tahun 2017 menggunakan dua metode *soft clustering*, yaitu metode *Fuzzy C-means* dan *Gaussian Mixture Model* (GMM), menunjukkan bahwa jumlah *cluster* pada hasil optimal adalah 2 dan 6. Jika dipertimbangkan dari konsistensi hasil validitas ketiga pengujian tersebut, dapat disimpulkan bahwa metode *Gaussian Mixture Model* (GMM) menghasilkan kelompok yang lebih baik dibandingkan dengan metode *Fuzzy C-means*. Kualitas kelompok yang diperoleh dipengaruhi oleh jenis data dan jumlah fitur data yang tersedia. Dengan demikian, penelitian ini membandingkan performa model antara *K-means*, *Fuzzy C-means*, dan *Gaussian Mixture Model* (GMM) yang bertujuan untuk mengidentifikasi metode terbaik dari ketiga model tersebut untuk data Indikator Kesehatan tahun 2021.

## 2. Metodologi

## 2.1 Bahan dan Data

Pada penelitian ini, kami menggunakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS), yang mencakup informasi tentang kondisi kesehatan di setiap provinsi di Indonesia pada tahun 2021. Data tersebut terdiri dari tujuh peubah bebas, di mana setiap variabel mengandung 34 observasi yang mewakili jumlah provinsi di Indonesia. Rincian mengenai variabel yang digunakan dapat ditemukan pada tabel berikut.

Tabel 1 Peubah yang digunakan

Peubah	Keterangan	Tipe Peubah
X1	Umur Harapan Hidup	Rasio (Numerik)
X2	Jumlah Puskesmas di Setiap Provinsi	Rasio (Numerik)
X3	Persentase balita sudah imunisasi dasar lengkap	Rasio (Numerik)
X4	Jumlah tenaga medis di setiap provinsi	Rasio (Numerik)
X5	Persentase penerima bantuan BPJS*	Rasio (Numerik)
X6	Persentase rumah tangga dengan sanitasi layak	Rasio (Numerik)
X7	Persentase rumah tangga dengan pelayanan sumber air minum layak	Rasio (Numerik)

\* Rasio dari total penerima bantuan BPJS dengan total penduduk pada provinsi tersebut

## 2.2 Tinjauan Pustaka

### 2.2.1 Clustering

*Clustering* adalah metode untuk menemukan sekelompok atau kumpulan objek data yang ditandai oleh kesamaan terbesar dalam *cluster* yang sama dan ketidaksetaraan terbesar diantara *cluster* yang berbeda. Dengan menggunakan metode ini, objek tersebut dikelompokkan bersama ke dalam satu atau beberapa *cluster* dengan tujuan agar memiliki kesamaan yang tinggi di antara satu sama lain (Metisen dan Sari 2015). Melalui *clustering*, kita dapat mengenali wilayah yang padat, mengungkap pola distribusi secara menyeluruh, dan menemukan hubungan menarik antar berbagai atribut data.

Terdapat dua pendekatan utama yang sering digunakan dalam metode *clustering*, yaitu pendekatan partisi dan pendekatan hierarki. Pendekatan partisi, seperti yang digunakan dalam algoritma *k-means clustering*, mengelompokkan beberapa objek dengan cara memilih objek-objek ke dalam *cluster* yang sudah ada. Sementara itu, pendekatan hierarki membuat pengelompokan objek dengan membentuk suatu hierarki berupa dendrogram (Nuryani dan Darwis 2021). Dua pendekatan ini adalah pendekatan utama yang digunakan dalam metode *clustering* untuk mengelompokkan data dengan cara yang berbeda sesuai dengan tujuan dan kebutuhan analisis.

### 2.2.2 Non Hierarchical Clustering

*Non Hierarchical Clustering* adalah teknik penggerombolan yang membagi data menjadi kelompok-kelompok yang tidak saling tumpang tindih (*non-overlapping groups*). Ini berbeda dari *hierarchical clustering* yang membuat hirarki dari kelompok-kelompok. Algoritma *Hierarchical Clustering* yang paling populer dan sederhana adalah *k-means clustering* (MacQueen, 1967).

### 2.2.3 K-Means

*K-Means* merupakan salah satu algoritma dalam *unsupervised learning* yang paling sederhana dan umum digunakan untuk menyelesaikan berbagai masalah pengelompokan. Tujuan utama dari *K-Means* adalah mengelompokkan data dengan cara menempatkan data ke dalam kelompok berdasarkan kedekatan dengan titik pusat *cluster* (centroid) terdekat (Darmansah dan Wardani 2021). Algoritma ini hanya dapat digunakan untuk atribut numerik karena didasarkan pada perhitungan jarak antar data untuk membagi data menjadi beberapa *cluster* (Hutagalung dan Sonata 2021). Langkah-langkah dari proses *clustering* dengan algoritma *K-Means* adalah sebagai berikut (Sani 2018):

1. Menentukan jumlah *cluster* yang diinginkan untuk melakukan pengelompokan data.
2. Data akan dialokasikan ke dalam *cluster* sesuai dengan jumlah yang ditentukan sebelumnya.
3. Nilai *centroid* (titik pusat) untuk setiap *cluster* ditentukan.
4. Jarak antara setiap data dengan *centroid cluster* yang terdekat dihitung menggunakan rumus *Euclidean*.
5. Hasil *clustering* ditampilkan berdasarkan jarak terendah dari hasil perhitungan pada langkah sebelumnya (keempat).
6. Jika hasil *clustering* belum sesuai, proses iterasi dilanjutkan dengan memperbarui nilai *centroid* pada langkah 3. Iterasi akan berhenti ketika hasil *clustering* sudah sama dengan iterasi sebelumnya yang menunjukkan konvergensi atau kestabilan dalam pembentukan *cluster*.

### 2.2.4 Fuzzy C-Means

*Fuzzy C-Means* (FCM) adalah algoritma penggerombolan non-hierarki yang memungkinkan titik data menjadi anggota dari beberapa klaster dengan tingkat keanggotaan yang bervariasi (Bezdek, 1981). Algoritma ini berdasarkan logika fuzzy dan merupakan penyempurnaan dari penggerombolan *k-means*.

Pada FCM, setiap titik data memiliki tingkat keanggotaan (*membership grade*) untuk setiap *cluster*. Tingkat keanggotaan menunjukkan seberapa besar titik data menjadi anggota dari berbagai *cluster*. Tingkat keanggotaan ini diperbarui secara iteratif menggunakan langkah-langkah berikut (Khan dan Ahman, 2004):

1. Inisialisasi tingkat keanggotaan secara acak
2. Menghitung *cluster prototypes* dengan mengambil rata-rata dari titik data, ditimbang oleh tingkat keanggotaan mereka terhadap *cluster* tersebut.
3. Menghitung jarak antara titik data dan *cluster prototypes*.
4. Memperbarui tingkat keanggotaan berdasarkan jarak dan fungsi keanggotaan fuzzy.

5. Mengulangi langkah 2-4 sampai konvergen atau jumlah iterasi maksimum tercapai.

FCM bertujuan untuk meminimalkan fungsi yang mewakili total bobot dalam ragam *cluster*. Fuzzifier m mengontrol penyamaran antar *cluster*. FCM konvergen ke minimum lokal dan sensitif terhadap inisialisasi dan *outlier*. Algoritma ini berfungsi baik pada *cluster* yang tumpang tindih (*overlapping clusters*) dan umumnya digunakan dalam pemrosesan citra (*image processing*), pengenalan pola, dan pengambilan keputusan. Kelebihan utamanya adalah fleksibilitas, kekokohan (*robustness*), dan kemampuan memodelkan ketidakpastian (Khan dan Ahman, 2004).

### 2.2.5 Gaussian Mixture Model

*Gaussian Mixture Model* (GMM) yang termasuk kedalam *Model-based clustering* mengoperasikan asumsi bahwa data dihasilkan oleh campuran distribusi peluang, dimana setiap komponen mewakili kelompok yang berbeda. Ini memungkinkan untuk memodelkan atau mengelompokkan individu dalam sebuah dataset ke dalam kelompok-kelompok yang sebelumnya tidak terdefinisi. Sebagai contoh, GMM memodelkan data sebagai kombinasi linier dari beberapa komponen Gaussian. (McLachlan dan Peel, 2000).

Algoritma GMM bekerja dengan memperkirakan parameter-parameter model (*mean*, matriks kovarians, dan bobot setiap komponen) menggunakan *Expectation-Maximization* (EM) agar *likelihood* data terhadap model dapat dimaksimalkan (Dempster *et al.*, 1977). Langkah-langkah algoritma GMM:

1. Inisialisasi parameter model secara acak
2. E-step: Hitung keanggotaan *cluster* data berdasarkan parameter saat ini
3. M-step: Perbarui parameter untuk memaksimalkan *likelihood*
4. Ulangi langkah 2-3 hingga konvergen

Penentuan jumlah *cluster* terbaik dapat dilakukan menggunakan model terbaik dengan pendekatan sebagai kriteria pemilihan model adalah *Bayes Information Criterion* (BIC) (Fraley dan Raftery, 2002). Model terbaik dipilih berdasarkan nilai BIC Terbesar.

Keunggulan GMM adalah mampu menangani *overlap cluster* dan *outlier* dengan baik. GMM telah banyak digunakan untuk *clustering* citra (*image*), teks, dan data *sequence*. Kekurangannya adalah hasil *clustering* bergantung pada inisialisasi dan jumlah komponen *Gaussian*.

### 2.2.6 Metode Evaluasi Model

*Clustering* bertujuan untuk mengelompokkan data yang memiliki kemiripan ke dalam beberapa *cluster*. Menentukan jumlah *cluster* yang optimal adalah tantangan utama dalam *clustering* sehingga evaluasi hasil *clustering* diperlukan.

Beberapa metode evaluasi *clustering* yang umum digunakan adalah *Davies-Bouldin Index* (DBI) dan *Silhouette Coefficient*. DBI mengukur rasio rerata *within-cluster scatter* dan *between-cluster separation* (Davies dan Bouldin, 1979). Nilai DBI yang lebih kecil mengindikasikan partisi *clustering* yang lebih baik.

*Silhouette Coefficient* mengukur seberapa baik objek cocok dalam kluster yang ditetapkan, dengan nilai -1 hingga 1 (Rousseeuw, 1987).

Kedua metode ini digunakan untuk menentukan jumlah *cluster* optimal secara objektif dari algoritma *clustering* seperti *K-means*, GMM, dan *hierarchical clustering*. Dengan melakukan evaluasi cluster, didapatkan model *clustering* terbaik sesuai karakteristik data. Evaluasi sangat penting dilakukan agar mendapatkan *clustering* yang valid.

## 2.3 Metode Penelitian

Pada penelitian ini, penulis menggunakan bantuan aplikasi perangkat lunak R. Adapun langkah-langkah untuk analisis data yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan eksplorasi data untuk melihat karakteristik data secara umum.
2. *Pre-processing* data dengan melakukan *cleaning* dan normalisasi. Adapun normalisasi yang diterapkan pada data yaitu normalisasi *min-max scaling*. Dengan normalisasi *min-max*, skala data diubah dari rentang aslinya sehingga semua nilai berada dalam kisaran 0 hingga 1 (Ambarwari et al. 2020).
3. Klasterisasi menggunakan Algoritma *K-Means* dengan langkah-langkah sebagai berikut (Sarbaini et al. 2022).
  - a. Menentukan *centroid* awal.
  - b. Menghitung jarak *Euclidean*.

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{jk})^2} \quad (1)$$

dengan  $x_{ij}$  adalah pusat kluster,  $c_{jk}$  adalah data indikator, dan  $d_{ik}$  adalah jarak setiap objek.

- c. Menghitung anggota kluster berdasarkan jarak terpendek.

$$\text{Min} \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{jk})^2} \quad (2)$$

- d. Menghitung *centroid* baru untuk iterasi berikutnya.

$$c = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

dengan  $p$  adalah banyaknya anggota kluster ke- $k$ .

- e. Ulangi langkah ketiga dan keempat, jika tidak terjadi perubahan anggota kluster maka iterasi berhenti dan hasil kluster didapatkan.
4. Klasterisasi menggunakan Algoritma *Fuzzy C-Means* dengan langkah-langkah sebagai berikut (Mailien et al. 2023):
  - a. Menentukan jumlah kluster ( $g$ ), parameter pembobot keanggotaan  $m'$ ,

matriks partisi awal  $U^{(0)}$ , dan tingkat akurasi  $\varepsilon_L$ .

- b. Menghitung pusat *cluster* ke- $l$  pada variabel ke- $j$  pada iterasi ke- $r$  menggunakan persamaan:

$$v_{lj}^{(r)} = \frac{\sum_{k=1}^n (\mu_{lk}^{(r)})^{m'} \cdot x_{kj}}{\sum_{k=1}^n (\mu_{lk}^{(r)})^{m'}} \quad (4)$$

- c. Menghitung jarak objek ke pusat kluster dan lakukan pembaruan elemen matriks partisi  $U^{(r)}$  menggunakan persamaan:

$$\mu_{lk}^{(r+1)} = \left[ \frac{g}{\sum_{i=1}^g \left( \frac{d_{lk}^{(r)}}{d_{ik}^{(r)}} \right)^{\frac{2}{(m'-1)}}} \right]^{-1} \quad (5)$$

di mana  $d_{lk}^{(r)}$  merupakan jarak objek ke- $k$  pada pusat kluster ke- $l$ .

- d. Menentukan konvergensi hasil pengelompokkan menggunakan persamaan:

$$\max_{l,k} \left| \mu_{lk}^{(r+1)} - \mu_{lk}^{(r)} \right| \leq \varepsilon_L \quad (6)$$

di mana  $\varepsilon_L$  merupakan tingkat akurasi yang ditentukan.

- e. Iterasi dihentikan ketika kriteria konvergensi terpenuhi. Jika belum terpenuhi, hitung kembali pusat kluster sampai kriteria konvergensi terpenuhi.
5. Klasterisasi menggunakan Algoritma *Gaussian Mixture Model* dengan langkah-langkah sebagai berikut:
- Inisialisasi model GMM dengan menentukan jumlah komponen *Gaussian* (kluster) yang diinginkan dan inisialisasi parameter seperti *mean*, *covariance*, dan proporsi kluster.
  - Menghitung probabilitas bahwa setiap titik data berasal dari masing-masing komponen *Gaussian* menggunakan model saat ini.
  - Perbarui parameter model berdasarkan probabilitas yang dihitung pada langkah ekspektasi.
  - Ulangi langkah ekspektasi dan maksimisasi secara bergantian sampai model konvergen.
6. Melakukan perbandingan hasil metode *K-Means*, *Fuzzy C-Means*, dan *Gaussian Mixture Model*. Ketiga metode clustering tersebut dibandingkan berdasarkan *Calinski-Harabasz Index* (CHI), *Davies-Bouldin Index* (DBI), dan *Silhouette Coefficient*.

7. Melakukan visualisasi hasil klaster model terbaik dan interpretasi *variable importance* pada setiap klaster.

## Daftar Pustaka

- [BPS] Badan Pusat Statistik. 2016. Potret awal tujuan pembangunan berkelanjutan (*sustainable development goals*) di Indonesia. Jakarta: Badan Pusat Statistik.
- [BPS] Badan Pusat Statistik. 2022. Indeks pembangunan manusia 2021. Jakarta: Badan Pusat Statistik.
- Amaluddin F. Muslim MA, Naba A. 2015. Klasifikasi Kendaraan Menggunakan Gaussian Mixture Model (GMM) dan Fuzzy Cluster Means (FCM). *Jurnal EECCIS*. 9(1) : 19-24.
- Ambarwari A, Adrian QJ, Herdiyeni Y. 2017. Analisis pengaruh data scaling terhadap performa algoritme machine learning untuk identifikasi tanaman. *JURNAL RESTI*. 1(3) : 117-122.
- Angelie AV. 2017. Segmentasi pelanggan menggunakan clustering K-Means dan model Rfm ( Studi Kasus : Pt . Bina Adidaya Surabaya ). *Its.*, siap terbit.
- Bezdek, J.C., Ehrlich, R., Full, W. 1984. FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2-3), 191-203.
- Darmansah, Wardani NW. 2021. Analisis persebaran penularan virus corona di provinsi jawa tengah menggunakan metode k-means clustering. *Jurnal Teknik Informatika dan Sistem Informasi*. 8(1) : 105-117.
- Davies, D. L., & Bouldin, D. W. 1979. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2), 224-227.
- Dempster, A.P., Laird, N.M., & Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39(1), 1-38.
- Fraley C, Raftery AE. 2002. Model based clustering, discriminant analysis, and density estimation. *J Amer Stat Assoc* 97.
- Hargono A, Waloejo C, Pandin MP, Choirunnisa Z. 2022. Penyuluhan pengolahan sanitasi air bersih untuk meningkatkan kesehatan masyarakat Desa Mengare, Gresik. *Abimanyu J Community Engagem.* 3(1):1–10. doi:10.26740/abi.v3n1.p1-10.
- Hutagalung J, Sonata F. 2021. Penerapan metode k-means untuk menganalisis minat sebuah nasabah asuransi. *Jurnal Media Informatika Budidarma*. 5(3) : 1187-1194



- Jacob DE, Sandjaya. 2018. Faktor faktor yang mempengaruhi kualitas hidup masyarakat Karubaga District Sub District Tolikara Provinsi Papua. *Jurnal Nasional Ilmu Kesehatan (JNIK)*. 1:1–16.
- Kementerian Kesehatan Republik Indonesia. 2021. Profil kesehatan Indonesia 2020. Jakarta.
- Khan, S.S., & Ahmad, A. 2004. Cluster center initialization algorithm for K-means clustering. *Pattern Recognition Letters*, 25(11), 1293-1302.
- MacQueen, J. B. 1967. Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press. pp. 281–297.
- Mailien B, Salma A, Syafriandi, Fitria D. 2023. Comparison K-Means and Fuzzy C-Means Methods to grouping human development index indicators in Indonesia. *UNP Journal of Statistics and Data Science*. 1(1): 23-30. doi:10.24036/ujsds/vol1-iss1/4.
- McLachlan, G.J. & Peel, D. 2000. *Finite mixture models*. Wiley.
- Metisen BM, Sari HL. 2015. Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada swalayan fadhila. *Jurnal Media Informa*. 11(2) : 110-118.
- Nugraha GS, Riyandari BA. 2020. Implementasi Fuzzy C-Means untuk pengelompokan daerah berdasarkan indikator kesehatan. *Jurnal Teknologi Informasi*. 4(1).
- Nuryani I, Darwis. 2021. Analisis clustering pada pengguna brand hp menggunakan metode k-means. *Prosiding Seminar Nasional Ilmu Komputer*. 1(1):190-211.
- Patel E, Kushwaha DS. 2020. Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. *Procedia Comput Sci*. 171(2019):158–167. doi:10.1016/j.procs.2020.04.017.
- Riyono J, Puspa SD, Pujiastuti CE. 2022. Simulasi clustering provinsi di Indonesia dalam penyebaran Covid-19 berdasarkan indikator kesehatan masyarakat menggunakan algoritma Gaussian Mixture Model . *Jurnal Matematika dan Pendidikan Matematika*. 5 (1) : 43-60.
- Rousseeuw, P. J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 53-65.
- Sani A. 2018. Penerapan metode k-means clustering pada perusahaan. *Jurnal Ilmiah Teknologi Informasi*. 353 : 1-7.

- Sarbaini, Saputri W, Nazaruddin, Muttakin F. 2022. Cluster analysis menggunakan algoritma Fuzzy K-Means untuk tingkat pengangguran di Provinsi Riau. *Jurnal Teknologi dan Manajemen Industri Terapan*. 1(2):78-84. doi:10.55826/tmit.v1i1.30.
- Saskia A, Jumiaty J. 2020. Efektivitas program sanitasi berbasis masyarakat dalam meningkatkan derajat kesehatan masyarakat. *J Manaj dan Ilmu Adm Publik*. 2(2):115–122. doi:10.24036/jmiap.v2i4.241.
- Suryani AS. 2020. Pembangunan air bersih dan sanitasi saat pandemi Covid-19. *Aspir J Masal Sos*. 11(2):199–214. doi:10.46807/aspirasi.v11i2.1757.
- Utami NH, Mubasyiroh R. 2019. Masalah gizi balita dan hubungannya dengan indeks pembangunan kesehatan masyarakat . *Penelitian Gizi Dan Makanan (The Journal of Nutrition and Food Research)*. 42(1): 1–10.
- Windia Ambarsari E, Dwitianti N, Selvia N, Nur Cholifah W, Dina Mardika P. 2023. Comparison approaches of the Fuzzy C-Means and Gaussian Mixture Model in clustering the welfare of the Indonesian people. *KnE Soc Sci*. 2023:16–22. doi:10.18502/kss.v8i9.13315.