

**ANALISIS KLASTER METODE PAM (*Partitioning Around Medoids*)-LITE
TERHADAP INDIKATOR KESEJAHTERAAN DAN KUALITAS HIDUP
NEGARA-NEGARA DI DUNIA**

(Skripsi)

Oleh

MAYA PUSPITASARI



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

**ANALISIS KLASTER METODE PAM (*Partioning Around Medoids*)-LITE
TERHADAP INDIKATOR KESEJAHTERAAN DAN KUALITAS HIDUP
NEGARA-NEGARA DI DUNIA**

**Oleh
MAYA PUSPITASARI**

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar

SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

ABSTRAK

Analisis Klaster Metode PAM (*Partitioning Around Medoids*)-*Lite* Terhadap Indikator Kesejahteraan dan Kualitas Hidup Negara-negara di Dunia

Oleh

Maya Puspitasari

Metode PAM (*Partitioning Around Medoids*) -*Lite* adalah suatu metode klasterisasi yang lebih cepat dan efisien dari algoritma PAM. PAM-*Lite* dikembangkan untuk mengatasi masalah toleransi terhadap pencilan dan data berukuran besar. Dalam penelitian ini, metode PAM-*Lite* digunakan untuk mengelompokkan negara-negara di dunia berdasarkan indikator kesejahteraan dan kualitas hidup. Berdasarkan hasil analisis klaster PAM-*Lite* dapat disimpulkan banyaknya klaster optimal adalah sebanyak 6 klaster dengan nilai *R-Square* sebesar 0,7858. Banyaknya anggota untuk setiap klaster adalah Klaster 1 terdapat 67 negara, Klaster 2 terdapat 69 negara, Klaster 3 terdapat 24 negara, Klaster 4 terdapat 37 negara, Klaster 5 terdapat 7 negara, dan Klaster 6 terdapat 13 negara.

Kata kunci : Data pencilan, Analisis klaster, PAM-*Lite*

ABSTRACT

Cluster Analysis Using PAM (Partitioning Around Medoids)-Lite Method on Welfare Indicators and Quality of Life in Countries Around The World

By

Maya Puspitasari

The PAM (Partitioning Around Medoids)-Lite method is a faster and more efficient clustering method than the PAM algorithm. PAM-Lite was developed to overcome the problem of outlier tolerance and large data size. In this study, the PAM-Lite method is used to cluster countries in the world based on welfare and quality of life indicators. Based on the results of the PAM-Lite cluster analysis, it can be concluded that the optimal number of clusters is 6 clusters with an R-Square value of 0.7858. The number of members for each cluster is obtained as follows : Cluster 1 is 67 countries, Cluster 2 is 69 countries, Cluster 3 is 24 countries, Cluster 4 is 37 countries, Cluster 5 is 7 countries, and Cluster 6 is 13 countries.

Kata kunci : Outliers, Cluster Analysis , *PAM-Lite*

Judul Skripsi : Analisis Klaster Metode PAM (*Partitioning Around Medoids*)-Lite Terhadap Indikator Kesejahteraan dan Kualitas Hidup Negara-negara di Dunia

Nama Mahasiswa : Maya Puspitasari

Nomor Pokok Mahasiswa : 2017031086

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam



MENYETUJUI,

1. Komisi Pembimbing

Dr. Khoirin Nisa, S.Si., M.Si.
NIP. 19740726 200003 2 001

Dra. Dorrah Aziz, M.Si.
NIP. 196101281988112001

Mengetahui,
Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

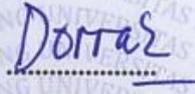
MENGESAHKAN

1. Tim Penguji

Ketua: **Dr. Khoirin Nisa, S.Si., M.Si.**

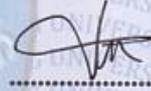


Sekretaris: **Dra. Dorrah Aziz, M.Si.**



Penguji

Bukan Pembimbing: **Drs. Nusyirwan, M.Si.**

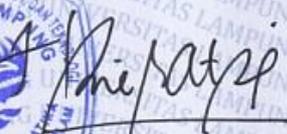


2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng Heri Satria, S.Si., M.Si.

NIP. 197111012005011002



Tanggal Lulus Ujian Skripsi : 28 Maret 2024

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Maya Puspitasari
Nomor Pokok Mahasiswa : 2017031086
Jurusan : Matematika
Judul Skripsi : Analisis Kluster Metode PAM (*Partitioning Around Medoids*)-Lite Terhadap Indikator Kesejahteraan dan Kualitas Hidup Negara-negara di Dunia

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua hasil tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 28 Maret 2024

Penulis



Maya Puspitasari

NPM. 2017031086

RIWAYAT HIDUP

Penulis bernama Maya Puspitasari lahir di Margototo pada 22 September 2001. Penulis merupakan anak kedua dari dua bersaudara dari pasangan Bapak Bowo dan Ibu Sukarmi.

Pada tahun 2020 penulis terdaftar sebagai mahasiswa S1 di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN. Selama menjadi mahasiswa, penulis aktif dalam organisasi Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) sebagai anggota biro Dana dan Usaha dan anggota Dinas ADKESMA Badan Eksekutif Mahasiswa (BEM FMIPA)

Pada Januari tahun 2023 penulis melaksanakan Kerja Praktik (KP) di Dinas Komunikasi dan Informatika Kota Metro. Pada Tahun yang sama di bulan Juli penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Gunung Rejo, Kecamatan Way Lima, Pesawaran selama 40 hari.

KATA INSPIRASI

“Sesungguhnya bersama kesulitan ada kemudahan”

(QS. Insyirah : 5)

“Bersabarlah engkau, janji Allah itu benar”

(QS. Ar-Rum : 60)

“Barang siapa yang mengerjakan kebaikan sekecil apapun, niscaya dia akan melihat (balasan)nya”

(QS. Al-Zalzalah : 7)

PERSEMBAHAN

Dengan mengucap puji dan syukur kepada Allah SWT yang telah memberikan petunjuk dan kemudahan untuk menyelesaikan perkuliahanku, kupersembahkan karya ini untuk :

Ayah dan Ibu tercinta yang selalu mendidik, mendoakan, mendukung, dan hal lainnya yang tidak dapat diungkapkan dengan kata-kata.

Abangku tersayang

Dosen pembimbing dan penguji yang sangat berjasa dan tidak lelah memberikan arahan serta masukan sehingga penulis dapat menyelesaikan skripsi ini.

Sahabat dan teman-temanku, terimakasih atas kebersamaan dan semangat yang selalu kalian berikan kepadaku.

Universitas Lampung.

SANWACANA

Puji dan syukur penulis ucapkan kepada Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Analisis Klaster Metode PAM (*Partitioning Around Medoids*)-Lite Terhadap Indikator Kesejahteraan dan Kualitas Hidup Negara-negara di Dunia”.

Penulis menyadari bahwa skripsi ini tidak akan terselesaikan dengan baik tanpa adanya arahan, bimbingan, serta kritik dan saran dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Ibu Dr. Khoirin Nisa, S.Si, M.Si., selaku dosen Pembimbing I serta dosen Pembimbing Akademik yang telah banyak memberikan bimbingan, kritik, saran, serta dukungan dalam penyelesaian skripsi ini.
2. Ibu Dra. Dorrah Aziz, M.Si. selaku dosen Pembimbing II atas bantuan dan bimbingan kepada penulis selama proses penyusunan skripsi ini berlangsung.
3. Bapak Drs. Nusyirwan, M.Si., selaku Dosen Pembahas yang telah memberikan kritik dan saran yang membangun selama proses penyusunan skripsi.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku ketua Jurusan Matematika.
5. Bapak Dr. Eng. Heri Satria, S.Si.,M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh dosen, staff, dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Kedua orang tua yang luar biasa, Bapak Bowo dan Mamak Sukarmi yang telah memberikan doa, dukungan dan kasih sayang yang tiada hentinya kepada penulis.

8. Abang Kurniawan Hadi Saputra, S.AP., M.H. dan Kakak Meutia Martha Shabrina, S.H., M.H. yang telah memotivasi dan selalu memberikan dukungan.
9. Fajar Agung Maryono Saputra, S.Mat. yang telah memberikan dukungan, kasih sayang dan bantuan sehingga skripsi ini dapat selesai
10. Teman-teman terbaik yang tidak bisa saya sebutkan satu persatu yang telah memberikan semangat dan dukungannya.
11. Teman-teman Matematika Angkatan 2020
12. Semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini.
13. Almamater tercinta Universitas Lampung.

Penulis menyadari bahwa penulisan skripsi ini masih banyak kekurangan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun guna penelitian selanjutnya.

Bandar Lampung, 28 Maret 2024

Penulis

Maya Puspitasari
NPM. 2017031086

DAFTAR ISI

	Halaman
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian.....	3
1.3 Manfaat Penelitian.....	3
II. TINJAUAN PUSTAKA	4
2.1 Matriks	4
2.1.1 Definisi Matriks	4
2.1.2 Determinan Matriks	4
2.1.3 Invers Matriks	5
2.1.4 Nilai Eigen dan Vektor Eigen	5
2.1.5 Matriks Korelasi.....	6
2.2 Analisis Kluster	7
2.2.1 Nilai Z (<i>Z-Score</i>).....	7
2.2.2 Ukuran Kemiripan (Jarak).....	8
2.2.3 Analisis Kluster K-Medoids	9
2.2.4 Analisis Kluster PAM- <i>Lite</i>	10
2.2.5 Indeks Evaluasi Analisis Kluster	12
2.3 Analisis Komponen Utama	14
2.4 Kesejahteraan dan Kualitas Hidup	15
III. METODOLOGI PENELITIAN	16
3.1 Waktu dan Tempat Penelitian	16
3.2 Data Penelitian	16
3.3 Metode Penelitian.....	16
IV. HASIL DAN PEMBAHASAN	20

4.1	Prapemrosesan Data	20
4.1.1	Standarisasi Data	20
4.1.2	Pengurangan Dimensi dengan Komponen Utama	20
4.1.3	Pemeriksaan Pencilan	22
4.2	Mencari Banyaknya Klaster (k) Optimal	22
4.3	Analisis Klaster PAM-Lite dengan k=6	24
4.3.1	Pilih n sampel, Setiap Sampel Memiliki Ukuran 40+2k	24
4.3.2	Jarak Euclid Setiap Pasangan Objek	25
4.3.3	Menghitung v_j untuk Objek j	26
4.3.4	Mengurutkan v_j , Pengklasteran, dan Menghitung Jumlah Jarak	28
4.3.5	Perbarui Medoid	29
4.3.6	Pengklasteran dan Menghitung Jumlah Jarak Baru	30
4.3.7	Menentukan Setiap Poin Terhadap Medoid Terdekat	32
4.3.8	Evaluasi Hasil Analisis Menggunakan R-Square dan Lebar <i>Silhouette</i>	35
V.	KESIMPULAN	38
	DAFTAR PUSTAKA	39
	LAMPIRAN	43

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Data terstandarisasi	20
Tabel 2. Korelasi antar variabel	20
Tabel 3. Nilai eigen	21
Tabel 4. Nilai R-Square dan <i>silhouette</i> pada tiap banyaknya klaster dengan 2 sampel	23
Tabel 5. Nilai R-Square dan <i>silhouette</i> pada tiap banyaknya klaster dengan 3 sampel	23
Tabel 6. Nilai R-Square dan <i>silhouette</i> pada tiap banyaknya klaster dengan 4 sampel	23
Tabel 7. Jarak Euclid antar objek pada sampel 1	26
Tabel 8. v_j untuk masing-masing objek.....	27
Tabel 9. Jarak Euclid antara medoid awal dan objek sampel 1	28
Tabel 10. Vektor klaster awal	29
Tabel 11. Jarak Euclid antara medoid baru dan objek sampel 1	30
Tabel 12. Vektor klaster baru.....	31
Tabel 13. Objek-objek yang dijadikan indeks medoid.....	31
Tabel 14. 15 jarak Euclid pertama	32
Tabel 15. Nilai rata-rata variabel antar klaster.....	33
Tabel 16. Karakteristik antar klaster	34
Tabel 17. Nilai rata-rata setiap variabel pada masing-masing klaster	36

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Diagram Alir Analisis Kluster PAM- <i>Lite</i>	18
Gambar 2. Diagram Alir Analisis Kluster PAM Terhadap Sampel	19
Gambar 3. Scatterplot komponen utama.....	21
Gambar 4. Boxplot pencilan	22
Gambar 5. Grafik nilai <i>R-Square</i> dan <i>silhouette</i> pada tiap banyaknya kluster	24
Gambar 6. Plot kluster data	33

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis kluster adalah teknik yang mengelompokkan data menjadi beberapa kelompok. Tujuan dari analisis kluster adalah untuk membagi data sehingga objek-objek dalam satu kelompok yang memiliki kesamaan karakteristik, sementara objek-objek di kelompok lain memiliki perbedaan (Arbelaitz *et al.*, 2013). Analisis kluster berperan penting dalam menghadapi kemajuan teknologi saat ini, contohnya pada konteks pemahaman tentang kesejahteraan dan kualitas hidup. Oleh karena itu, mengembangkan metode-metode baru dalam analisis kluster penting sekali agar dapat menghasilkan hasil yang akurat dan signifikan.

Menurut Wibowo dkk. (2022), algoritma pengelompokan dirancang untuk mengidentifikasi struktur dasar data dan menggunakan hubungan yang terdeteksi dalam struktur tersebut untuk mengelompokkan objek ke dalam kelompok yang berbeda. Algoritma pengelompokan yang sering digunakan, yaitu K-Means dan K-Medoids. Kedua metode ini berperan penting dalam mengelompokkan data, meskipun memiliki karakteristik yang berbeda. K-Means digunakan untuk menentukan kelompok data dengan menghitung centroid, yang merupakan titik tengah berdasarkan rata-rata dari data dalam kelompok tersebut (Yedla *et al.*, 2010). Metode ini sering digunakan, terutama ketika data cenderung berdistribusi secara normal. Namun, K-Means memiliki kelemahan saat menangani data yang memiliki nilai-nilai pencilan (*outliers*), karena centroid dapat terpengaruh oleh nilai-nilai ekstrem ini. Sehingga, dalam kasus ini K-Means mungkin menghasilkan kelompok-kelompok yang tidak optimal. K-Medoids atau dapat disebut PAM (*Partitioning Around Medoids*) dapat mengatasi masalah ini, karena PAM lebih tahan terhadap *outliers* (Prahara *et al.*, 2020). PAM tidak menghitung *centroid*, tetapi mencari titik-titik data yang paling representatif dalam kelompok (medoid). Dengan demikian,

PAM lebih stabil terhadap nilai-nilai ekstrem dalam data sehingga cocok untuk situasi di mana *outliers* dapat memengaruhi hasil analisis secara signifikan. Namun, terdapat varian yang lebih ringan dan efisien dari PAM yang disebut dengan PAM-*Lite*.

Menurut Olukanmi *et al.* (2019), PAM-*Lite* lebih unggul daripada PAM dalam hal toleransi terhadap *outliers* dan data yang berukuran besar. PAM-*Lite* dengan pendekatan yang lebih sederhana dan lebih efisien dalam waktu komputasi seringkali mampu mengatasi kehadiran *outliers* dengan lebih baik. Hal ini terutama disebabkan oleh cara PAM-*Lite* memilih medoid awal dan pengurangan iterasi yang lebih sedikit dalam upaya mengoptimalkan pengelompokan. Ketika memiliki data yang mengandung pencilan yang dapat mempengaruhi hasil kluster, PAM-*Lite* dapat menjadi pilihan yang lebih efektif dan terpercaya untuk mempertahankan kualitas pengelompokan tanpa mengorbankan waktu komputasi yang berlebihan.

Penelitian mengenai analisis kluster dengan indikator kesejahteraan dan kualitas hidup sebelumnya pernah dilakukan Žmuk (2015), membahas analisis kluster metode ward tentang peningkatan harapan hidup manusia dan kualitas hidup sebagai faktor penting, hasilnya menunjukkan tiga kelompok negara: negara Uni Eropa lama, negara Uni Eropa baru, dan negara non-Uni Eropa mengindikasikan hubungan antara kualitas hidup dan perkembangan ekonomi. Wibowo dkk. (2022), juga melakukan analisis kluster K-Medoids atau PAM berdasarkan data COVID-19 dari 219 negara. Hasilnya, negara-negara tersebut dikelompokkan menjadi lima kluster berdasarkan jumlah kasus dan kematian kumulatif, kasus positif per kapita, dan *case fatality rate*. Ini membantu memahami variasi dampak pandemi di seluruh dunia.

Kesejahteraan dan kualitas hidup adalah faktor penting dalam kehidupan manusia. Indikator kesejahteraan seperti *contributing family workers*, *employment*, dan *life expectancy* memiliki dampak signifikan pada kualitas hidup individu dan masyarakat. Berdasarkan penelitian-penelitian di atas, maka dalam penelitian ini

peneliti ingin melakukan analisis kluster dengan metode *PAM-Lite* pada data indikator kesejahteraan dan kualitas hidup.

1.2 Tujuan Penelitian

Tujuan penelitian ini adalah menerapkan metode analisis kluster *PAM-Lite* dalam mengelompokkan negara-negara di dunia berdasarkan indikator kesejahteraan dan kualitas hidup.

1.3 Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi terhadap pengembangan pengetahuan matematika dalam analisis kluster.
2. Menerapkan metode analisis kluster *PAM-Lite* pada indikator kesejahteraan dan kualitas hidup untuk merumuskan kebijakan yang lebih efisien dalam meningkatkan kualitas hidup masyarakat.

II. TINJAUAN PUSTAKA

2.1 Matriks

2.1.1 Definisi Matriks

Menurut Anton & Rorres (1994), matriks adalah susunan bilangan, simbol, atau ekspresi yang disusun dalam baris dan kolom. Bilangan-bilangan dalam matriks disebut entri atau elemen matriks. Baris adalah susunan bilangan yang mendatar (horizontal), dan kolom adalah susunan bilangan yang tegak (vertikal). Matriks A dengan m kolom dan n baris disebut matriks mxn , dan dapat ditulis sebagai berikut:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

2.1.2 Determinan matriks

Determinan matriks didefinisikan sebagai selisih antara perkalian elemen-elemen pada diagonal utama dengan perkalian elemen-elemen pada diagonal sekunder. Determinan matriks hanya dapat digunakan pada matriks persegi. Determinan dari matriks A dapat dituliskan $\det(A)$ atau $|A|$ (Anton & Rorres, 1994). Adapun determinan pada matriks A sebagai berikut:

$$\det(A) = |A| = \sum_{j=1}^n a_{ij}A_{ij}$$

Dengan :

n : ukuran matriks

a_{ij} : elemen matriks pada baris ke- i dan kolom ke- j

A_{ij} : kofaktor dari elemen a_{ij} .

2.1.3 Invers Matriks

Menurut Anton & Rorres (1994), invers matriks adalah matriks baru yang merupakan kebalikan dari matriks asal. Jika \mathbf{A} adalah matriks persegi, dan jika terdapat matriks \mathbf{B} yang ukurannya sama sedemikian rupa sehingga $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, dengan \mathbf{I} melambangkan matriks identitas $n \times n$, dan perkalian yang dilakukan merupakan perkalian matriks yang umum. Jika hubungan tersebut berlaku, maka matriks \mathbf{B} disebut sebagai invers dari matriks \mathbf{A} , dan diberi lambang \mathbf{A}^{-1} . Adapun invers matriks \mathbf{A} sebagai berikut:

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \text{Adj } \mathbf{A}$$

2.1.4 Nilai Eigen dan Vektor Eigen

Matriks persegi dengan ukuran $n \times n$ memiliki nilai dan vektor karakteristik. Istilah yang lebih umum dari nilai karakteristik adalah nilai eigen dan vektor karakteristik adalah vektor eigen. Definisi dari nilai eigen dan vektor eigen (Pratiwi dkk., 2017) sebagai berikut:

Jika \mathbf{A} adalah matriks $n \times n$, maka suatu vektor tak nol \mathbf{v} pada \mathbf{R}^n disebut vektor eigen dari \mathbf{A} jika \mathbf{Av} adalah sebuah kelipatan skalar dari \mathbf{v} yaitu

$$\mathbf{Av} = \lambda \mathbf{v}$$

untuk sebarang skalar λ . Skalar λ disebut nilai eigen dari \mathbf{A} , dan \mathbf{v} disebut sebagai vektor eigen dari \mathbf{A} yang bersesuaian dengan λ .

Untuk memperoleh nilai eigen dari sebuah matriks \mathbf{A} berukuran $n \times n$, dapat ditulis sebagai $\mathbf{A}\mathbf{v} = \lambda\mathbf{I}\mathbf{v}$ atau ekuivalen dengan

$$(\lambda\mathbf{I} - \mathbf{A})\mathbf{v} = 0 \quad (1)$$

agar λ dapat menjadi nilai eigen, harus terdapat paling sedikit satu solusi tak nol dari Persamaan 1. Namun, Persamaan 1 memiliki solusi tak nol jika dan hanya jika

$$\det(\lambda\mathbf{I} - \mathbf{A}) = 0 \quad (2)$$

Persamaan 2 disebut persamaan karakteristik matriks \mathbf{A} . Skalar-skalar yang memenuhi persamaan tersebut adalah nilai-nilai eigen matriks \mathbf{A} .

2.1.5 Matriks Korelasi

Matriks korelasi adalah matriks yang menunjukkan koefisien korelasi berpasangan antar variabel. Koefisien korelasi adalah angka yang mengukur kekuatan dan arah hubungan linier antara dua variabel. Nilainya berkisar antara -1 hingga 1 , dimana -1 menunjukkan korelasi negatif sempurna, 1 menunjukkan korelasi positif sempurna, dan 0 menunjukkan tidak ada korelasi. Matriks korelasi dapat mengidentifikasi variabel mana yang terkait satu sama lain dan seberapa kuatnya (Graybill, 1983). Misalkan \mathbf{y} adalah vector acak berukuran $n \times 1$, dan matriks kovarian yang dinotasikan dengan \mathbf{V} . Maka, korelasi matriks dari \mathbf{y} dinotasikan dengan $\mathbf{R} = [\rho_{ij}]$ dimana ρ_{ij} dapat ditulis sebagai berikut:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

Matriks korelasi dari \mathbf{A} berukuran $m \times n$ dapat ditulis sebagai berikut:

$$R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & 1 \end{bmatrix}$$

2.2 Analisis Kluster

Analisis kluster adalah teknik multivariat yang bertujuan untuk mengelompokkan objek sehingga keragaman dalam kluster menjadi sekecil mungkin, sementara keragaman antar kluster menjadi sebesar mungkin (Härdle & Simar, 2019). Analisis kluster didasarkan pada jarak, dan jarak yang paling umum digunakan adalah jarak Euclidean untuk mengukur kemiripan antar objek (Suhaeni dkk., 2018). Ini adalah sebuah teknik statistik yang digunakan untuk mengelompokkan objek atau data ke dalam kelompok-kelompok (kluster) berdasarkan kesamaan karakteristik atau atribut tertentu. Tujuannya adalah untuk mengidentifikasi struktur dalam data yang memungkinkan objek dalam satu kluster memiliki tingkat kemiripan yang tinggi satu sama lain, sementara secara signifikan berbeda dari objek dalam kluster lainnya.

2.2.1 Nilai Z (Z-Score)

Z-Score merupakan metode statistik yang digunakan untuk mengukur hubungan antara suatu skor dengan rata-rata dalam suatu kelompok skor (Kumar & Rao, 2015). Nilai *Z-Score* didefinisikan sebagai berikut:

$$Z = \frac{X - \mu}{\sigma} \quad (2.1)$$

Keterangan:

Z : nilai standar

X : nilai yang akan distandarisasi

μ : rata-rata distribusi

σ : simpangan baku (standar deviasi) distribusi

2.2.2 Ukuran Kemiripan (Jarak)

Analisis kluster adalah proses pengelompokan data berdasarkan sejauh mana data-data tersebut mirip atau dekat satu sama lain. Hal ini menghasilkan kelompok-kelompok (kluster) di mana data-data dalam satu kelompok memiliki tingkat kemiripan yang tinggi, sedangkan antar kelompok memiliki tingkat kemiripan yang rendah.

Pada penelitian ini jarak Euclidean digunakan sebagai alat ukur kemiripan, yang didefinisikan sebagai berikut:

$$d(i, j) = d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.2)$$

Keterangan:

d_{ij} : jarak antara objek ke-i pada objek ke-j

x_{ik} : data dari objek ke-i pada variabel ke-k

x_{jk} : data dari objek ke-j pada variabel ke-k

(Suwanda *et al.*, 2019)

Pada analisis kluster PAM, penggunaan jarak Euclidean memungkinkan fleksibilitas dalam definisi "kedekatan" antara elemen-elemen data (Van der Laan *et al.*, 2003). Hal ini berarti bahwa dengan PAM, dapat menggunakan jarak Euclidean untuk menentukan seberapa dekat atau kemiripan dua objek. Sementara pada analisis kluster K-Means, penggunaan jarak Euclidean juga umum dan digunakan untuk mengukur jarak antara setiap objek data dan cluster centroid (Yadav & Sharma, 2013).

2.2.3 Analisis Kluster K-Medoids

K-Medoids atau PAM (*Partitioning Around Medoids*) yang diajukan oleh Kaufman dan Rousseeuw, adalah metode yang menggambarkan kluster yang dibentuk dengan menggunakan medoid sebagai pusat kluster. Algoritma ini memilih satu objek dalam kumpulan objek sebagai representasi dari sebuah kluster, dan objek ini disebut sebagai medoid. PAM mewarisi karakteristik dari algoritma pengelompokan K-Medoids. Di samping itu, PAM dilengkapi dengan mekanisme pertukaran medoids yang dapat meningkatkan hasil pengelompokan. PAM terbukti lebih tahan terhadap gangguan dan *outliers* daripada K-means karena medoids dalam PAM lebih sedikit dipengaruhi oleh *outliers* (Pande *et al.*, 2012).

Langkah-langkah analisis kluster PAM (Gultom *et al.*, 2018) sebagai berikut:

- a. Lakukan perhitungan jarak antara setiap pasangan objek berdasarkan ukuran jarak yang dipilih.
- b. Menghitung nilai v_j untuk objek j dengan menggunakan persamaan:

$$v_j = \sum_{i=1}^n \frac{d_{ij}}{\sum_{l=1}^n d_{il}}; \quad j = 1, 2, 3, \dots, n \quad (2.3)$$

- c. Urutkan nilai v_j dari yang terkecil hingga terbesar dan pilih k objek dengan nilai v_j terkecil sebagai medoid awal.
- d. Hasilkan hasil awal dari pengklasteran dengan menentukan objek yang paling dekat dengan medoid.
- e. Menghitung jumlah jarak dari setiap objek ke medoidnya.
- f. Temukan medoid baru dari setiap klaster dengan cara meminimalkan jumlah jarak di dalam klaster tersebut dan perbarui medoid dari setiap klaster dengan medoid yang baru.
- g. Tentukan setiap objek ke medoid terdekat dan peroleh hasil pengklasteran.
- h. Lakukan perhitungan jumlah jarak dari semua objek ke medoidnya.

2.2.4 Analisis Klaster PAM-Lite

Algoritma PAM-Lite merupakan versi cepat dari algoritma pengelompokan PAM. Algoritma PAM-Lite unggul dalam hal efisiensi jika dibandingkan dengan PAM, bahkan mampu menghasilkan pengelompokan yang lebih baik daripada PAM. Algoritma PAM-Lite menggunakan pendekatan inferensia yang sama dengan K-Means-Lite. Perbedaan antara kedua algoritma ini adalah bahwa dalam PAM-Lite, sampel dan himpunan gabungan pusat klaster yang diperoleh dari sampel tersebut dikelompokkan menggunakan PAM, berbeda dengan K-Means-Lite yang didasarkan pada K-Means (Olukanmi *et al.*, 2019). Namun, algoritma PAM-Lite didasarkan pada paradigma K-Means-Lite.

2.2.4.1 Paradigma K-Means-Lite

Algoritma K-Means-Lite dikenalkan sebagai solusi untuk mengatasi masalah skalabilitas yang kurang baik pada algoritma K-Means. Menurut Olukanmi *et al.* (2019) cara kerja K-Means-Lite berdasarkan perluasan teorema limit pusat

(CLT). CLT memberikan dasar untuk membuat kesimpulan tentang suatu populasi menggunakan sampel acak. Ini menjelaskan bahwa jika diberikan populasi P dengan rata-rata μ dan deviasi standar σ , maka distribusi dari rata-rata sampel \bar{X}_i ($i = 1, \dots, n$) dimana n merupakan sampel yang masing-masing berukuran s yang dipilih secara acak dari populasi tersebut yang cenderung akan berdistribusi normal yang memiliki rata-rata μ dan deviasi standar σ/\sqrt{s} , saat $n, s \rightarrow \infty$. Pertama, jika suatu populasi besar akan dipartisi, dan jumlah kluster $k = 1$, maka μ adalah centroid. Dengan kata lain, CLT menunjukkan bahwa μ adalah centroid dari populasi, dapat diestimasi dengan cara mengelompokkan setiap dari n sampel berukuran yang sama s ke dalam $k = 1$ kluster, yang menghasilkan rata-rata \bar{X}_i sebagai centroid dari setiap sampel dan menggabungkan n rata-rata ini menjadi satu dataset, yang ketika digabungkan menjadi $k = 1$ cluster menghasilkan centroid c yang mendekati μ , saat s dan n meningkat. Oleh karena itu, metode yang berfokus pada inferensia ini merupakan alternatif yang lebih efisien untuk melakukan perhitungan terhadap centroid.

2.2.4.2 Algoritma PAM-Lite

Pendekatan inferensia pada sub-bab 2.1.4.1 untuk k-means, juga dapat diterapkan pada pengelompokan PAM. Dasar dari kedua metode ini adalah bahwa medoids dan centroid keduanya adalah perkiraan pusat kluster. Oleh karena itu, algoritma PAM-Lite menggunakan pendekatan inferensia yang identik dengan K-Means-Lite. Perbedaan antara kedua algoritma tersebut adalah bahwa dalam PAM-Lite, sampel-sampel dan kumpulan pusat kluster yang diperoleh dari sampel-sampel tersebut dikelompokkan menggunakan PAM, berbeda dengan K-Means-Lite yang berdasarkan pada K-Means. Algoritma PAM-Lite diuraikan sebagai berikut:

1. Pilih n sampel, masing-masing dengan ukuran s dari dataset P
2. Jalankan PAM pada setiap sampel dan simpan medoids kluster dalam P'

3. Jalankan PAM pada P' untuk mendapatkan himpunan medoids C
4. Tetapkan setiap titik dalam P ke medoid terdekat dalam C .

Untuk pemilihan ukuran sampel s , mengikuti nilai-nilai yang telah divalidasi secara empiris untuk *K-Means-Lite*, yaitu mengikuti ukuran sampel algoritma CLARA (Kaufman & Rousseeuw, 2008), yaitu memilih n sampel dan masing-masing sampel berukuran $40 + 2k$ (Olukanmi *et al.*, 2018).

2.2.5 Indeks Evaluasi Analisis Kluster

2.2.5.1 Nilai R-Square

Nilai *R-square* dalam analisis kluster digunakan untuk menentukan apakah terdapat perbedaan yang signifikan di antara objek-objek dalam kelompok-kelompok yang berbeda dan bahwa objek-objek dalam kelompok yang sama memiliki kemiripan yang tinggi. Jika RS sama dengan nol, maka tidak ada perbedaan antara kelompok-kelompok tersebut. Di sisi lain, jika RS sama dengan satu, maka pemisahan kluster optimal (Rujasiri & Chomtee, 2009). Untuk menghitung nilai *R-Square* dapat menggunakan persamaan 2.4.

$$RS = \frac{SS_B}{SS_T} = \frac{SS_T - SS_W}{SS_T} \quad (2.4)$$

$$RS = \frac{\left\{ \sum_{j=1}^p \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right\} - \left\{ \sum_{k=1}^c \sum_{j=1}^p \sum_{i=1}^{n_c} (x_{ijk} - \bar{x}_{jk})^2 \right\}}{\left\{ \sum_{j=1}^p \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right\}}$$

Keterangan:

x_{ij} : objek ke- i pada variabel j

\bar{x}_j : rata-rata pada variabel j

x_{ijk} : objek ke- i pada klaster k pada variabel j

\bar{x}_{jk} : rata-rata pada variabel j pada klaster k

p : banyaknya variabel

c : banyaknya klaster

n_c : banyaknya objek pada klaster c

2.2.5.2 Lebar *Silhouette*

Definisi lebar *Silhouette* menurut Rousseeuw (1987) dapat diuraikan sebagai berikut:

Jika sebuah objek yang menjadi pusat/fokus dan termasuk dalam klaster A. Kelompok lain yang tidak termasuk objek ini sebagai klaster C. Nilai $a(i)$ didefinisikan sebagai rata-rata perbedaan antara objek i dengan semua objek lain yang ada dalam kelompok A. Sedangkan nilai $c(i, C)$ adalah rata-rata perbedaan antara objek i dan semua objek yang terdapat dalam kelompok C.

$$b(i) = \min_{C \neq A} c(i, C)$$

Lebar *Silhouette* $s(i)$, didefinisikan sebagai berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2.5)$$

Lebar *Silhouette* $s(i)$, berkisar antara -1 hingga 1. Nilai yang mendekati 1 mengindikasikan bahwa objek tersebut jauh lebih mirip dengan objek-objek dalam kelompok yang sama daripada dengan objek-objek dari kelompok terdekat lainnya, menunjukkan bahwa klasifikasinya sangat baik. Jika $s(i)$ mendekati 0, ini menunjukkan bahwa klasifikasi objek fokus tersebut diragukan, dan objek tersebut berada dalam posisi antara dua kelompok. Nilai $s(i)$ yang mendekati -1 menunjukkan kesalahan klasifikasi yang jelas. Oleh karena itu, rata-rata lebar

siluet di seluruh kelompok memberikan penilaian tentang seberapa baik kualitas kelompok tersebut, atau rata-rata dari seluruh sampel juga dapat digunakan sebagai indikator validitas keseluruhan klasifikasi (Lengyel & Botta-Dukát, 2019).

2.3 Analisis Komponen Utama

Analisis Komponen Utama (AKU) adalah metode matematika yang digunakan untuk mengurangi dimensi dataset, meningkatkan kemampuan interpretasi, dan secara bersamaan mengurangi kehilangan informasi (Jolliffe & Cadima, 2016). AKU bekerja dengan merubah variabel asli ke dalam sebuah himpunan variabel baru yang tidak memiliki hubungan satu sama lain, yang disebut sebagai komponen utama. Komponen utama ini dibentuk melalui kombinasi linier dari fitur asli, dan mereka mampu menangkap sebanyak mungkin informasi dari data dengan mengurangi kerugian informasi sebisa mungkin. Adapun beberapa tahapan dalam melakukan analisis komponen utama diantaranya yaitu: menghitung matriks korelasi, menghitung nilai eigen, menghitung vektor eigen, dan menghitung komponen utama. Persamaan analisis komponen utama adalah sebagai berikut:

$$KU_1 = \alpha_{11}x_1 + \alpha_{21}x_2 + \alpha_{31}x_3 + \dots + \alpha_{p1}x_p$$

$$KU_2 = \alpha_{12}x_1 + \alpha_{22}x_2 + \alpha_{32}x_3 + \dots + \alpha_{p2}x_p$$

$$\vdots \quad \vdots$$

$$KU_j = \alpha_{1j}x_1 + \alpha_{2j}x_2 + \alpha_{3j}x_3 + \dots + \alpha_{pj}x_p$$

Keterangan:

KU_j = komponen utama ke-j

α_{pj} = koefisien untuk variabel ke-p dalam komponen utama ke-j

x_p = nilai variabel ke-p dari data asli

2.4 Kesejahteraan dan Kualitas Hidup

Menurut Ruggeri *et al.* (2020), kesejahteraan dan kualitas hidup merupakan konsep multidimensi yang dapat dinilai dengan menggunakan berbagai indikator dan ukuran. Kesejahteraan dan kualitas hidup suatu negara sering kali diukur melalui sejumlah indikator penting, termasuk *contributing family workers*, *employment*, dan *life expectancy*.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2023/2024 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung

3.2 Data Penelitian

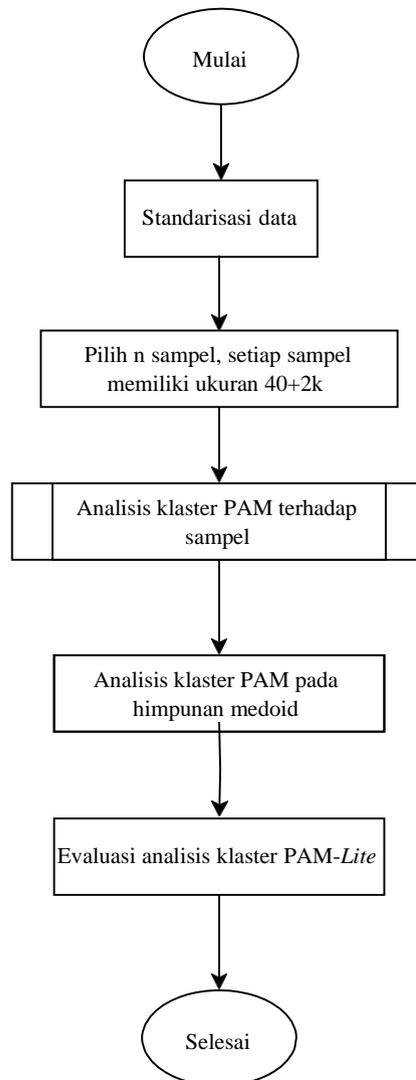
Data yang digunakan pada penelitian ini adalah data sekunder mengenai data kesejahteraan dan kualitas hidup dengan variabel negara, *contributing family workers*, *employment*, dan *life expectancy* yang berjumlah 217 data dan diambil dari situs databank.worldbank.org.

3.3 Metode Penelitian

Adapun langkah-langkah penelitian ini sebagai berikut:

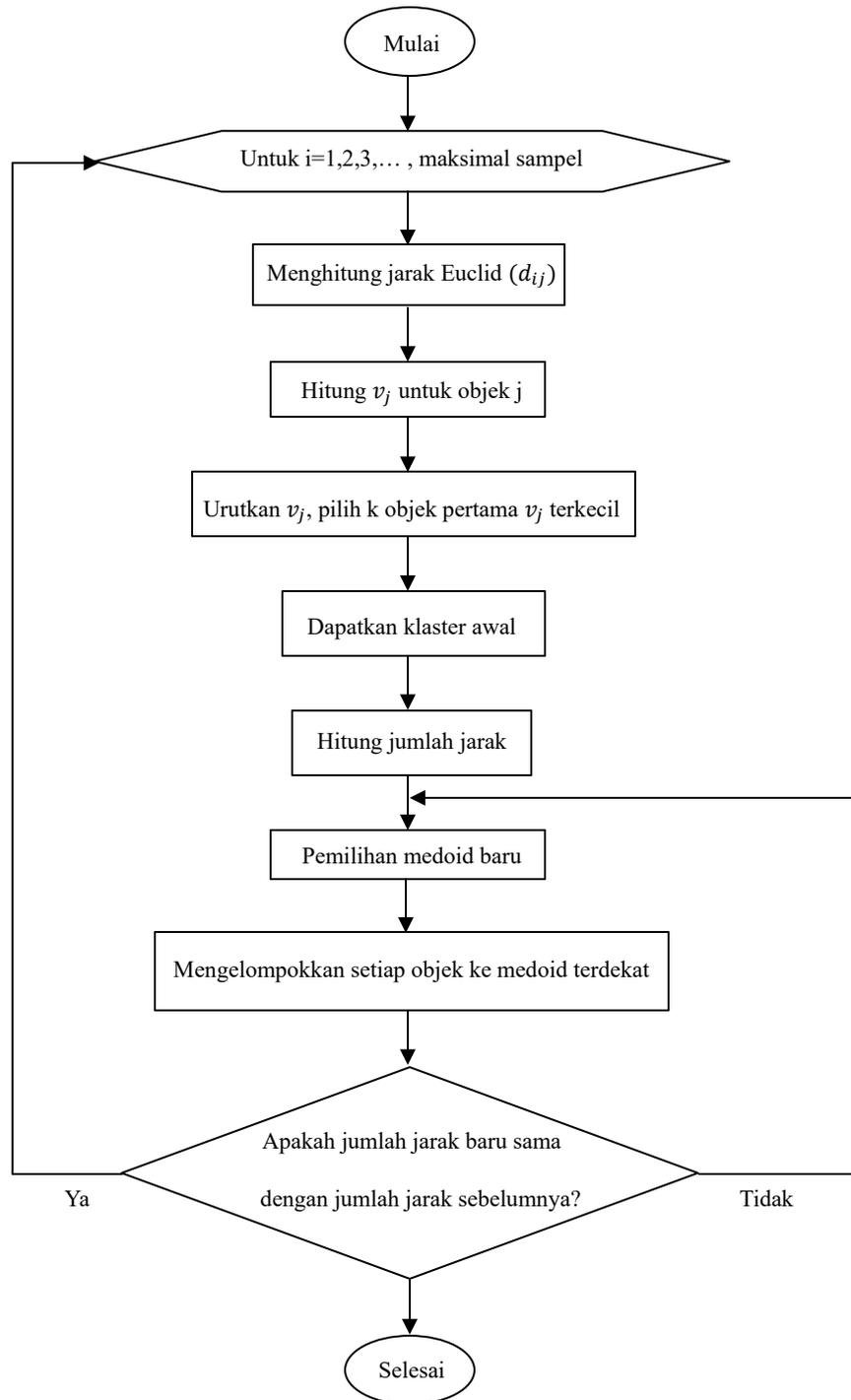
1. Melakukan standarisasi data
2. Pengurangan dimensi dengan komponen utama
3. Memilih n sampel, dengan setiap sampel memiliki ukuran $40 + 2k$
4. Melakukan analisis kluster PAM terhadap sampel
 - a. Menghitung jarak Euclid setiap pasangan objek
 - b. Hitung v_j untuk objek j
 - c. Urutkan v_j , pilih k objek pertama v_j terkecil sebagai inisial medoid
 - d. Dapatkan hasil kluster awal dengan menetapkan setiap poin objek ke medoid terdekat
 - e. Hitung jumlah jarak dari semua objek terhadap medoidnya
 - f. Pemilihan medoid baru untuk setiap kluster yang meminimalkan total jarak

- g. Mengelompokan setiap objek ke medoid terdekat
 - h. Hitung jumlah jarak baru dari semua objek terhadap medoid
5. Melakukan analisis kluster PAM pada himpunan medoid yang dihasilkan oleh langkah-langkat di atas
 6. Melakukan evaluasi analisis kluster PAM-*Lite* yang disajikan dalam bentuk evaluasi hasil analisis menggunakan *R-square* dan lebar *silhouette*.



Gambar 1. Diagram Alir Analisis Kluster PAM-Lite

Analisis kluster PAM terhadap sampel



Gambar 2. Diagram Alir Analisis Kluster PAM Terhadap Sampel

V. KESIMPULAN

Berdasarkan hasil penelitian, diperoleh bahwa penerapan analisis kluster PAM-*Lite* terhadap data indikator kesejahteraan dan kualitas hidup negara-negara di dunia menghasilkan jumlah kluster optimal sebesar 6 kluster dengan nilai *R-Square* sebesar 0,7858. Banyaknya anggota untuk setiap kluster adalah kluster 1 terdapat 67 negara, kluster 2 terdapat 69 negara, kluster 3 terdapat 24 negara, kluster 4 terdapat 37 negara, kluster 5 terdapat 7 negara, dan kluster 6 terdapat 13 negara.

DAFTAR PUSTAKA

- Anton, H., & Rorres, C. 1994. *Elementary linear algebra: applications version*. John Wiley & Sons.
- Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J. M., & Perona, I. 2013. An extensive comparative study of cluster validity indices. *Pattern Recognition*, **46**(1), 243–256.
- Graybill, F. A. 1983. *Matrices with Applications in Statistics*. Taipei, Taiwan: Second Edition, Wadsworth Publishing company.
- Gultom, S., Sriadhi, S., Martiano, M., & Simarmata, J. 2018. Comparison analysis of K-means and K-medoid with Euclidean distance algorithm, Manhattan distance, and Chebyshev distance for big data clustering. *IOP Conference Series: Materials Science and Engineering*, **420**(1), 120-139.
- Härdle, W. K., & Simar, L. 2019. *Applied Multivariate Statistical Analysis 5th Edition*. New York: Springer.
- Hodge, V. J., & Austin, J. 2004. A survey of outlier detection methodologies. *Artificial Intelligence Review*, **22**(2), 85-126.
- Jolliffe, I. T., & Cadima, J. 2016. Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, **374**(2065), 20150202.

- Kaufman, L., & Rousseeuw, P. J. 1990. Clustering large applications (Program CLARA). In *Wiley Series in Probability and Statistics*. 126-163. John Wiley & Sons, Ltd.
- Kumar, M. N., & Rao, V. S. 2015. A New Methodology for Estimating Internal Credit Risk and Bankruptcy Prediction under Basel II Regime. *Computational Economics*, **46**(1), 83–102.
- Lengyel, A., & Botta-Dukát, Z. 2019. Silhouette width using generalized mean—A flexible method for assessing clustering efficiency. *Ecology and Evolution*, **9**(23), 13231–13243.
- Mann, A. K., & Kaur, N. 2013. Review paper on clustering techniques. *Global Journal of Computer Science and Technology*, **13**(5), 43-47.
- Olukanmi, P. O., Nelwamondo, F., & Marwala, T. 2019. PAM-lite: Fast and accurate k-medoids clustering for massive datasets. *Proceedings-2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa, SAUPEC/RobMech/PRASA 2019, typically 5*, 200-204.
- Olukanmi, P., Nelwamondo, F., & Marwala, T. 2018. k-means-lite: real time clustering for large datasets. *5th International Conference on Soft Computing and Machine Intelligence*, 54-59.
- Pande, S. R., Sambare, S. S., & Thakre, V. M. 2012. Data clustering using data mining techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, **1**(8), 494-499.

- Prahara, A., Ismi, D. P., & Azhari, A. 2020. Parallelization of Partitioning Around Medoids (PAM) in K-Medoids Clustering on GPU. *Knowledge Engineering and Data Science*, **3**(1), 40-49.
- Pratiwi, Y. E., Kiftiah, M., & Ramadhani, E. W. 2017. Penentuan Nilai Eigen dan Vektor Eigen Matriks Interval Menggunakan Metode Pangkat. *Buletin Ilmiah Math. Stat. dan Terapannya*, **6**(2), 17 – 26.
- Rousseeuw, P. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Computational and Applied Mathematics*, **20**, 53–65.
- Ruggeri, K., Garcia-Garzon, E., Maguire, Á., Matz, S., & Huppert, F. A. 2020. Well-being is more than happiness and life satisfaction: a multidimensional analysis of 21 countries. *Health and quality of life outcomes*, **18**(1), 1-16.
- Rujasiri, P., & Chomtee, B. 2009. Comparison of clustering techniques for cluster analysis. *Kasetsart Journal Natural Science*, **43**(2), 378-388.
- Starmans, M. P., van der Voort, S. R., Tovar, J. M., Veenland, J. F., Klein, S., & Niessen, W. J. 2020. *Starmans, M. P. A., van der Voort, S. R., Castillo Tovar, J. M., Veenland, J. F., Klein, S. Radiomics. In Handbook of Medical Image Computing and Computer Assisted Intervention*. Amsterdam, The Netherlands: Elsevier.
- Suhaeni, C., Kurnia, A., & Ristiyanti. 2018. Perbandingan Hasil Pengelompokan menggunakan Analisis Cluster Berhierarchy, K-Means Cluster, dan Cluster Ensemble (Studi Kasus Data Indikator Pelayanan Kesehatan Ibu Hamil). *Jurnal Media Infotama*, **14**(1), 31-38.

- Suwanda, R., Syahputra, Z., & Zamzami, M. 2019. Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K. *Journal of Physics*, **1566**(1). IOP Publishing.
- Van der Laan, M., Pollard, K., & Bryan, J. 2003. A new partitioning around medoids algorithm. *Journal of Statistical Computation and Simulation*, **73**(8), 575–584.
- Wibowo, R. A., Nisa, K., Venelia, H., & Warsono. 2022. Robust Clustering of COVID-19 Pandemic Worldwide. *Barekeng : Jurnal Ilmu Matematika dan Terapan*, **16**(2), 687-694.
- Yadav, J., & Sharma, M. 2013. A Review of K-mean Algorithm. *International Journal of Engineering Trends and Technology*, **4**(7), 2972-2976.
- Yedla, M., Pathakota, S. R., & Srinivasa, T. M. 2010. Enhancing K-means clustering algorithm with improved initial center. *International Journal of computer science and information technologies*, **1**(2), 121-125.
- Žmuk, B. 2015. Quality of life indicators in selected European countries: Statistical hierarchical cluster analysis approach. *Croatian Review of Economic, Business and Social Statistics*, **1**(1-2), 42-54.