PENERAPAN ALGORITMA *K-MEDOIDS* PADA INDIKATOR TUJUAN PEMBANGUNAN BERKELANJUTAN DI INDONESIA DENGAN EVALUASI *DUNN INDEX*

ALSA MUARTI
NIM H1091211066

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

PENERAPAN ALGORITMA *K-MEDOIDS* PADA INDIKATOR TUJUAN PEMBANGUNAN BERKELANJUTAN DI INDONESIA DENGAN EVALUASI *DUNN INDEX*

ALSA MUARTI
NIM H1091211066

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

LEMBAR PENGESAHAN

Judul Tugas Akhir	:	Penerapan Algoritma *K-Medoids* pada Indikator Tujuan Pembangunan Berkelanjutan di Indonesia dengan Evaluasi *Dunn Index*
Nama Mahasiswa	:	Alsa Muarti
NIM	:	H1091211066
Jurusan/Program Studi	:	Matematika/Statistika
Tanggal Lulus	:	11 Februari 2025
SK Pembimbing	:	No. 2640/UN22.8/TD.06/2024/Tanggal 2 September 2024
SK Penguji	:	No. 278/UN22.8/TD.06/2025/Tanggal 3 Februari 2025

Dosen Pembimbing

Pembimbing I	Pembimbing II
Shantika Martha, M.Si.	Hendra Perdana, M.Sc.
NIP. 198403082008122003	NIP. 198810102019031020

Dosen Penguji

Ketua Penguji	Anggota Penguji
Nurfitri Imro'ah, M.Si.	Dr. Evy Sulistianingsih, M.Sc.
NIP. 198907182019032021	NIP. 198502172008122006
Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Shantika Martha, M.Si.	Hendra Perdana, M.Sc.
NIP. 198403082008122003	NIP. 198810102019031020

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, M.Si.
NIP. 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama : Alsa Muarti

NIM : H1091211066

Program Studi/ Jurusan : Statistika/Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyataan ini untuk digunakan sebagaimana mestinya.

Pontianak, 11 Februari 2025

Alsa Muarti
NIM. H1091211066

Penerapan Algoritma *K-Medoids* pada Indikator Tujuan Pembangunan Berkelanjutan di Indonesia dengan Evaluasi *Dunn Index*

Abstrak

Pembangunan berkelanjutan adalah pembangunan yang memperhatikan berbagai aspek dan tidak hanya fokus pada kebutuhan sekarang tetapi juga mempertimbangkan kepentingan generasi mendatang. Pembangunan berkelanjutan dapat diukur dari aspek ekonomi, sosial dan lingkungan yang diwujudkan dalam ukuran keadilan, kenyamanan, dan kelestarian. Pengelompokkan indikator tujuan pembangunan berkelanjutan menjadi langkah penting untuk memastikan kesesuaian kebijakan dengan kondisi suatu wilayah. Tujuan penelitian ini adalah mengelompokkan provinsi di Indonesia berdasarkan indikator tujuan pembangunan berkelanjutan serta melakukan evaluasi tehadap *cluster* yang terbentuk. Analisis *cluster* merupakan suatu analisis multivariat yang digunakan untuk mengelompokkan data berdasarkan kesamaan antara satu data dengan data lainnya ke dalam kelompok atau *cluster*. Algoritma *K-Medoids* merupakan salah satu analisis *cluster* untuk mengelompokkan data dengan menggunakan *medoids* sebagai pusat setiap *cluster*. Selanjutnya, evaluasi dalam analisis *cluster* penting dilakukan agar seluruh provinsi di Indonesia dapat dikelompokkan dengan tepat. *Dunn Index* (DI) merupakan salah satu metode evaluasi pada *cluster* yang digunakan untuk menilai kualitas *cluster* serta membantu menentukan jumlah *cluster* optimum berdasarkan rasio antara jarak minimum antar *cluster* dan diameter maksimum dalam *cluster*. Data yang digunakan pada penelitian ini adalah 10 indikator tujuan pembangunan berkelanjutan untuk setiap provinsi di Indonesia tahun 2023. Hasil penelitian menunjukkan jumlah *cluster* terbaik sebanyak 3 *cluster* dengan nilai DI sebesar 0,28 yang merupakan nilai DI terbesar dari jumlah *cluster* lainnya. Pada *cluster* 1 menunjukkan kondisi pembangunan yang tinggi dengan anggota 14 provinsi, *cluster* 2 beranggotakan 19 provinsi dengan kondisi pembangunan yang sedang, dan *cluster* 3 dengan kondisi pembangunan rendah yang beranggotakan 1 provinsi, yaitu Provinsi Papua.

Kata kunci: *Z-Score*, Jarak *Euclidean,* Multikolinearitas, PAM

Application of K-Medoids Algorithm on Sustainable Development Goal Indicators in Indonesia with Dunn Index Evaluation

Abstract

Sustainable development is development that takes into account various aspects and not only focuses on current needs but also considers the interests of future generations. Sustainable development can be measured from economic, social and environmental aspects which are realized in measures of justice, comfort and sustainability. Grouping indicators of sustainable development goals is an important step to ensure the suitability of policies with the conditions of a region. The purpose of this study is to group provinces in Indonesia based on indicators of sustainable development goals and evaluate the clusters formed. Cluster analysis is a multivariate analysis used to group data based on similarities between one data and other data into groups or clusters. The K-Medoids algorithm is one of the cluster analysis for data grouping using medoids as the center of each cluster. Furthermore, evaluation in cluster analysis is important so that all provinces in Indonesia can be grouped appropriately. Dunn Index (DI) is one of the evaluation methods on clusters used to assess the quality of clusters and help determine the optimum number of clusters based on the ratio between the minimum distance between clusters and the maximum diameter within the cluster. The data used in this study are ten indicators of sustainable development goals for each province in Indonesia in 2023. The results showed the best number of clusters as many as 3 clusters with a DI value of 0.28 which is the largest DI value of the number of other clusters. Cluster 1 shows high development conditions with 14 members, cluster 2 consists of 19 provinces with moderate development conditions, and cluster 3 with low development conditions consisting of 1 province, namely Papua Province.

Keywords: Z-Score, Euclidean Distance, Multicollinearity, PAM

PRAKATA

Assalamu'alaikum Warahmatullahi Wabarakatuh

Alhamdulillahirabbil'alamin, puji syukur ke hadirat Allah SWT atas segala rahmat, karunia dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "**Penerapan Algoritma *K-Medoids* pada Indikator Tujuan Pembangunan Berkelanjutan di Indonesia dengan Evaluasi *Dunn Index***". Tugas akhir ini dibuat dengan tujuan untuk memenuhi syarat memperoleh gelar Sarjana Statistika bagi mahasiswa Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Pontianak.

Selama proses penyusunan tugas akhir ini, penulis menyadari telah banyak mendapatkan bimbingan, bantuan, dukungan dan motivasi dari berbagai pihak. Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah terlibat dalam penyusunan tugas akhir ini, terutama kepada:

1. Kedua orang tua yang penulis cintai dan sayangi yaitu Ibunda Elpina dan Ayahanda Tasripin (Alm) yang selalu memberikan semangat, dukungan, dan doa yang sangat berarti dan tiada hentinya untuk penulis.
2. Ibu Shantika Martha, M.Si. selaku dosen pembimbing pertama yang telah memberikan bimbingan dan arahan dalam penulisan tugas akhir ini.
3. Bapak Hendra Perdana, M.Sc. selaku dosen pembimbing kedua yang telah memberikan bimbingan dan arahan dalam penulisan tugas akhir ini.
4. Ibu Nurfitri Imro'ah, M.Si. selaku dosen penguji pertama yang telah memberikan saran dan masukan kepada penulis.
5. Ibu Dr. Evy Sulistianingsih, M.Sc. selaku dosen penguji kedua yang telah memberikan saran dan masukan kepada penulis.
6. Ibu Meliana Pasaribu, M.Sc. selaku dosen pembimbing akademik yang telah memberikan bimbingan selama perkuliahan.
7. *Community Development & Outreaching* Universitas Tanjungpura selaku pengelola beasiswa parsial yang telah memberikan bantuan berupa biaya hidup selama perkuliahan.

8. Abang, kakak, dan adik yang penulis sayangi yaitu Edwin, Widiarti, Lidiarti, dan Bagus Alfayet telah memberikan semangat, dukungan, dan doanya untuk penulis.
9. Sahabat seperjuangan yang penulis sayangi yang sama-sama menimba ilmu di Program Studi Statistika yaitu Eri Sindia yang telah membantu dan memberikan semangat kepada penulis dalam menyelesaikan tugas akhir ini.
10. Sahabat-sahabat yang penulis sayangi yaitu Wanda, Artina, Tiara, Hani, Annisa dan Aimi yang selalu mendoakan dan memotivasi penulis dalam menyelesaikan tugas akhir ini.
11. Rekan-rekan Mahasiswa Statistika Universitas Tanjungpura, khususnya "Exposs" yang namanya tidak dapat penulis sebutkan satu persatu yang telah memberikan dukungan kepada penulis.
12. Seluruh pihak yang tidak dapat disebutkan satu persatu yang telah memberikan doa, bantuan, dan semangat dalam penyelesaian tugas akhir ini.

Akhir kata semoga tugas akhir ini bermanfaat bagi semua pihak yang membutuhkan.

Pontianak, 11 Februari 2025

Alsa Muarti

DAFTAR ISI

LEMBAR PENGESAHAN ..... **ii**
PERNYATAAN INTEGRITAS AKADEMIK ..... **iii**
Abstrak ..... **iv**
Abstract ..... **v**
PRAKATA ..... **vi**
DAFTAR ISI ..... **viii**
DAFTAR TABEL ..... **x**
DAFTAR GAMBAR ..... **xi**
DAFTAR SIMBOL ..... **xii**
DAFTAR LAMPIRAN ..... **xiii**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan 4
1.4 Batasan Masalah 4
1.5 Tinjauan Pustaka 4
1.6 Metodologi Penelitian 5
BAB II LANDASAN TEORI **9**
2.1 Pembangunan Berkelanjutan 9
2.2 Analisis Multivariat 10
2.3 Analisis *Cluster* 10
2.3.1 Metode Hierarki (*Hierarchical Methods*) 11
2.3.2 Metode *Non*-Hierarki (*Non-Hierarchical Methods*) 12
2.4 Uji Multikolinearitas 13
2.5 Standarisasi *Z-score* 13
2.6 Ukuran Jarak 14
BAB III EVALUASI *CLUSTER* **15**
3.1 Algoritma *K-Medoids* 15
3.2 *Dunn Index* 16

BAB IV HASIL DAN PEMBAHASAN 18
4.1 Statistik Deskriptif 18
4.2 Uji Multikolinearitas 19
4.3 Standarisasi Data 20
4.4 Pengelompokkan Menggunakan Algoritma *K-Medoids* 21
4.5 Evaluasi *Cluster* Menggunakan *Dunn Index* 23
4.6 Intepretasi Profil *Cluster* 26
BAB V PENUTUP 30
5.1 Kesimpulan 30
5.2 Saran 30
DAFTAR PUSTAKA 32
LAMPIRAN 36

DAFTAR TABEL

Tabel 4.1 Variabel Penelitian 18
Tabel 4.2 Statistik Deskriptif 19
Tabel 4.3 Nilai *Variance Inflation Factor* (VIF) 20
Tabel 4.4 Pengelompokan Provinsi di Indonesia dengan 2 *Cluster* 23
Tabel 4.5 Nilai *Dunn Index* 25
Tabel 4.6 Pengelompokkan Provinsi di Indonesia pada 3 *Cluster* 26
Tabel 4.7 Rata-Rata Variabel Tujuan Pembangunan Berkelanjutan Tiap *Cluster* 26

DAFTAR GAMBAR

Gambar 1.1 *Flowchart* Penelitian..... 8
Gambar 4.1 Peta Hasil Pengelompokan Provinsi di Indonesia 28

DAFTAR SIMBOL

p	:	banyak variabel
R_l^2	:	koefisien determinasi antara X_l dengan variabel independen lainnya
VIF_l	:	*Variance Inflation Factor* (VIF) untuk variabel independen ke-l, dengan $l = 1,2, \dots, p$
n	:	banyak objek yang diamati
x_i	:	objek ke-i, dengan $i = 1, 2, \dots, n$
x_j	:	objek ke-j, dengan $j = 1, 2, \dots, n$
$Z_{i,l}$	:	nilai standarisasi objek ke-i variabel ke-l, dengan $i = 1, 2, \dots, n$
$x_{i,l}$	:	nilai objek ke-i variabel ke-l
X_l	:	nilai variabel ke-l, dengan $l = 1,2, \dots, p$
μ_l	:	rata-rata variabel ke-l, dengan $l = 1, 2, \dots, p$
σ_l	:	standar deviasi variabel ke-l, dengan $l = 1, 2, \dots, p$
$d(x_i, x_j)$	:	jarak *euclidean* antara objek ke-i dan objek ke-j
K	:	banyak *cluster* yang terbentuk
m_k	:	*medoids* pada *cluster* ke-k
S	:	selisih total jarak
DI_K	:	*Dunn Index* pada jumlah *cluster*
$d(c_k, c_m)$	:	jarak terdekat antar *cluster* ke-k dan *cluster* ke-m
c_k	:	*cluster* ke-k, dengan $k = 1,2, \dots, K$
c_m	:	*cluster* ke-m, dengan $m = k + 1, \dots, K$
x_{i,c_k}	:	objek ke-i yang berada pada c_k, dengan $i = 1,2, \dots, n$
x_{j,c_m}	:	objek ke-j yang berada pada c_m, dengan $j = 1,2, \dots, n$
$diam(c_k)$	:	jarak maksimal antar anggota dalam *cluster* ke-k

DAFTAR LAMPIRAN

Lampiran I Data Tujuan Pembangunan Berkelanjutan 2023 36
Lampiran II Standarisasi Data 37
Lampiran III *Euclidean Distance K-Medoids Clustering* Pada Iterasi 1 38
Lampiran IV *Euclidean Distance K-Medoids Clustering* Pada Iterasi 2 39
Lampiran V Nilai Jarak Dalam Satu *Cluster* pada 2 *Cluster* 40
Lampiran VI Nilai Jarak Antar *Cluster* Pada 2 *cluster* 42

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pembangunan berkelanjutan telah menjadi fokus utama bagi banyak negara di dunia. Hal ini disebabkan oleh kemajuan pesat dalam teknologi, industrialisasi, dan urbanisasi yang terjadi di berbagai belahan dunia seringkali membawa dampak yang tidak hanya bersifat positif, tetapi juga negatif terhadap lingkungan dan sosial masyarakat. Dampak negatif yang dirasakan, seperti kerusakan lingkungan, ketidaksetaraan sosial, dan penurunan kualitas hidup, sehingga menjadi peringatan bagi para pemimpin untuk merumuskan strategi pembangunan yang lebih berwawasan ke depan dan berkelanjutan (Prasetyo, 2023).

Menurut Brundtland Report dari PBB (1987) di dalam Setiani dan Hakim (2015) pembangunan berkelanjutan adalah proses pembangunan yang dilakukan dengan memperhatikan kebutuhan generasi sekarang tanpa mengorbankan kemampuan generasi mendatang dalam memenuhi kebutuhan. Oleh karena itu, setiap kebijakan pembangunan harus mempertimbangkan tiga aspek utama, yaitu aspek lingkungan, ekonomi, dan sosial. Aspek lingkungan dapat digambarkan dengan variabel akses terhadap layanan sumber air minum layak dan akses terhadap layanan sanitasi layak (Setianingtias, Baiquni, dan Kurniawan, 2015). Sedangkan aspek sosial diukur dari persentase penduduk miskin, umur harapan hidup, rata-rata lama sekolah, indeks pembangunan gender dan rumah tangga mempunyai akses listrik (Bangun, 2020), serta aspek ekonomi yang diukur dari pengeluaran per kapita, rumah tangga dengan status kepemilikan rumah milik dan tingkat pengangguran terbuka (Setianingtias, Baiquni, dan Kurniawan, 2015). Tujuan utama pembangunan berkelanjutan adalah menciptakan kesejahteraan jangka panjang tanpa merusak alam dan mengabaikan keadilan sosial.

Namun, penerapan konsep pembangunan berkelanjutan seringkali terhambat oleh berbagai tantangan. Adanya pola strategi pembangunan konvensional menjadi tantangan bagi para pengambil kebijakan dalam mengatasi degradasi lingkungan akibat eksploitasi yang berlebihan terhadap sumber daya alam. Kerusakan

lingkungan yang semakin parah akan berujung pada masalah baru pada lingkungan itu sendiri, seperti deforestasi, peningkatan polusi udara, perubahan iklim, dan menurunnya kandungan pangan. Hal ini menunjukkan bahwa pembangunan yang berfokus pada keuntungan ekonomi semata telah menciptakan dampak negatif yang merusak daya dukung bumi (Febriana, Diartho, dan Istiyani, 2019). Sementara itu, ketimpangan sosial dan ekonomi juga menjadi masalah besar, di mana sebagian pendapatan suatu negara hanya dapat dinikmati segelintir pihak dan pihak lainnya hanya mendapatkan porsi yang kecil atau malah tidak mendapatkannya. Ketimpangan dalam pembangunan adalah cikal bakal terbentuknya kemiskinan serta berbagai macam masalah sosial yang penanganannya membutuhkan waktu yang tidak sebentar. Oleh karena itu pembangunan yang merata hendaknya harus segera terealisasikan (Andiny dan Mandasari, 2017).

Salah satu masalah sosial yang terus menjadi hambatan dalam pembangunan berkelanjutan adalah kemiskinan. Meskipun angka kemiskinan nasional mengalami penurunan 0,18% pada tahun 2023 dibandingkan tahun sebelumnya, wilayah Indonesia bagian timur, khususnya Papua dan Nusa Tenggara Timur masih mencatatkan angka kemiskinan yang sangat tinggi, yaitu masing-masing sebesar 26,03% dan 19,96%. Angka ini jauh diatas DKI Jakarta yang hanya sebesar 4,44% (Badan Pusat Statistik, 2023). Ketimpangan ini menunjukkan bahwa meskipun ada perbaikan secara nasional, kemiskinan tetap menjadi masalah besar di beberapa wilayah. Hal ini mengindikasikan adanya kesenjangan dalam pencapaian pembangunan. Kesenjangan akibat kemiskinan ini dapat membatasi akses masyarakat terhadap pendidikan, layanan kesehatan, air bersih, dan pekerjaan yang layak, sehingga menciptakan lingkaran kemiskinan yang sulit diputus (Ibrahim, 2017).

Indonesia menghendaki pembangunan yang baik, yaitu pembangunan yang tidak hanya berfokus pada kebutuhan saat ini, tetapi juga mempertimbangkan keberlanjutan untuk generasi mendatang. Untuk mewujudkan target ini, diperlukan strategi yang tepat sasaran dan disesuaikan dengan karakteristik setiap wilayah. Kondisi topografi dan lingkungan yang berbeda menyebabkan permasalahan pembangunan yang dihadapi pada masing-masing daerah berbeda-beda. Kondisi

pembangunan pada tiap-tiap daerah penting untuk diketahui agar membantu pemerintah dalam menentukan kebijakan. Salah satu pendekatan yang dapat digunakan adalah dengan melakukan identifikasi dan pengelompokan daerah melalui analisis *cluster*. Metode ini memungkinkan segmentasi wilayah berdasarkan karakteristik dan masalah yang serupa, sehingga dapat membantu dalam merancang strategi pembangunan yang lebih terarah dan sesuai dengan kebutuhan masing-masing wilayah. Algoritma *K-Medoids* adalah metode analisis *cluster* yang dapat mengatasi *outlier* atau pencilan. Metode ini menitik beratkan pada homogenitas di dalam satu kelompok.

Permasalahan dalam analisis *cluster* ini adalah belum diketahuinya jumlah *cluster* yang baik digunakan, sehingga membuat objek belum terklaster dengan baik dan tepat. Evaluasi *cluster* digunakan untuk menilai kualitas *cluster* serta membantu menentukan jumlah *cluster* optimum. Berbagai metode seperti *Elbow, Silhouette, Dunn Index* (DI)*, Calinski Harabasz, Pseudo-F Statistic,* dan *Davies Bouldin Index* dapat digunakan untuk mengevaluasi *cluster*. *Dunn Index* merupakan metode evaluasi *cluster* yang dapat mengukur kualitas pengelompokan dengan memaksimalkan jarak antar *cluster* dan meminimalkan jarak dalam *cluster*. Semakin tinggi nilai *Dunn Index*, maka semakin optimal jumlah *cluster* yang dihasilkan (Malikhatin, Rusgiyono, dan Maruddani, 2021). Dengan menggabungkan evaluasi kohesi dan separasi antar *cluster, Dunn Index* memberikan gambaran menyeluruh tentang efektivitas suatu pengelompokan (Hidayatullah, Martha, dan Aprizkiyandari, 2024). Metode *Dunn Index* diharapkan dapat membantu pengklasteran 34 provinsi di Indonesia dengan cepat, sehingga masalah dapat teraratasi dari kebijakan yang tepat sasaran dan efektif.

1.2 Rumusan Masalah

Berdasarkan penjelasan latar belakang yang diuraikan, rumusan masalah dalam penelitian ini adalah bagaimana pembentukan *cluster* menggunakan algoritma *K-Medoids* pada indikator tujuan pembangunan berkelanjutan di Indonesia dan penentuan jumlah optimum *cluster* dengan evaluasi *Dunn index*.

1.3 Tujuan

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini adalah membentuk hasil *cluster* menggunakan algoritma *K-Medoids* pada indikator tujuan pembangunan berkelanjutan di Indonesia dan menentukan jumlah *cluster* optimum menggunakan evaluasi *Dunn Index*.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini hanya membahas data dari indikator tujuan pembangunan berkelanjutan di Indonesia tahun 2023 yang bersumber dari Badan Pusat Statistik (BPS). Data indikator pembangunan berkelanjutan yang digunakan yaitu meliputi umur harapan hidup, rata-rata lama sekolah, pengeluaran per kapita, tingkat pengangguran terbuka, indeks pembangunan gender, persentase penduduk miskin, rumah tangga dengan status kepemilikan rumah milik, rumah tangga yang memiliki akses terhadap sanitasi yang layak, rumah tangga yang mempunyai akses listrik, dan rumah tangga yang memiliki akses terhadap sumber air minum layak. Pengelompokan indikator tujuan pembangunan berkelanjutan pada 34 provinsi di Indonesia dengan analisis *cluster non-hierarchical* menggunakan *K-Medoids* dengan evaluasi *Dunn Index* dan pembentuk *cluster* optimum yang diuji berjumlah 2, 3, 4, dan 5 *cluster*.

1.5 Tinjauan Pustaka

Rudianto dan Wijayanto (2023) melakukan pengelompokan 34 provinsi di Indonesia berdasarkan aspek yang terdapat dalam Indeks Demokrasi Indonesia (IDI) tahun 2021. Metode *clustering* yang digunakan dalam penelitian ini yaitu *K-Means* dan *K-Medoids*. Pemilihan jumlah *cluster* yang optimal menggunakan *Dunn Index*, *Silhouette*, *Calinski-Harabasz*, dan *Davies-Bouldin Index*. Variabel yang digunakan adalah aspek kebebasan, aspek kesetaraan, dan aspek kapasitas lembaga demokrasi, serta diperoleh 5 *cluster* pada IDI. Hasil yang diperoleh menggunakan metode *Dunn Index*, *Silhouette*, dan *Calinski-Harabasz* menyatakan bahwa algoritma *K-Medoids* lebih baik dibandingkan *K-Means*.

Analisis perbandingan metode *K-Medoids* dan *Fuzzy C-Means* dilakukan oleh Sihotang, Bekti, dan Jatipaningrum (2024). Penelitian ini bertujuan

mengelompokkan kabupaten/kota di Provinsi Sumatera Utara berdasarkan jenis kekerasan terhadap anak. Variabel yang digunakan, yaitu kekerasan secara fisik, psikis, seksual, dan penelantar. Berdasarkan penelitian ini, metode *K-Medoids* menghasilkan 4 *cluster* dan pada metode *Fuzzy C-Means* menghasilkan 2 *cluster*. Hasil perbandingan uji validasi *Silhouette* dan *Dunn Index* menunjukkan bahwa metode *clustering* terbaik terdapat pada algoritma *K-Medoids* dengan nilai *Silhouette* sebesar 0,53 dan *Dunn Index* sebesar 0,22.

Luthfi dan Wijayanto (2021) melakukan penelitian pengelompokan Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2019. Variabel yang digunakan terdiri dari angka harapan hidup, angka melek huruf, rata-rata lamanya sekolah, dan pengeluaran per kapita. Metode yang digunakan metode *Hierarchical*, *K-Means*, dan *K-Medoids Clustering*. Hasil yang diperoleh terdapat 5 *cluster* yang paling optimum menggunakan evaluasi *Dunn Index* dari metode *K-Medoids* dengan nilai sebesar 0,03.

Wira, Budianto, dan Wiguna (2019) melakukan analisis *clustering* untuk mengetahui pola pemilihan program studi mahasiswa baru di Universitas Kanjuruhan Malang tahun 2018. Variabel yang digunakan dalam penelitian ini nilai ujian, asal sekolah, dan program studi yang dipilih. Hasil yang diperoleh terdapat 3 *cluster* yang paling optimum menggunakan algoritma *K-Medoids* dengan evaluasi *Silhouette Coefficient* dan objek yang digunakan sebanyak 5. Hasil *Clustering* menunjukkan bahwa mahasiswa baru yang berasal dari sekolah SMA dan SMK dengan nilai ujian diatas 70 mengambil jurusan Teknik Informatika, sedangkan mahasiswa baru yang berasal dari sekolah SMK dengan nilai ujian dibawah 70 dan SMA dengan nilai ujian dibawah 50 mengambil jurusan Sistem Informasi, dan sisanya mengambil jurusan selain Teknik Informatika dan Sistem Informasi.

1.6 Metodologi Penelitian

Alat bantu analisis dalam penelitian ini adalah *software* R dan Microsoft Excel yang digunakan untuk membantu dalam pengelompokan provinsi di Indonesia berdasarkan indikator tujuan pembangunan berkelanjutan dengan menggunakan algoritma *K-Medoids*. Penelitian ini menggunakan data sekunder yang diperoleh

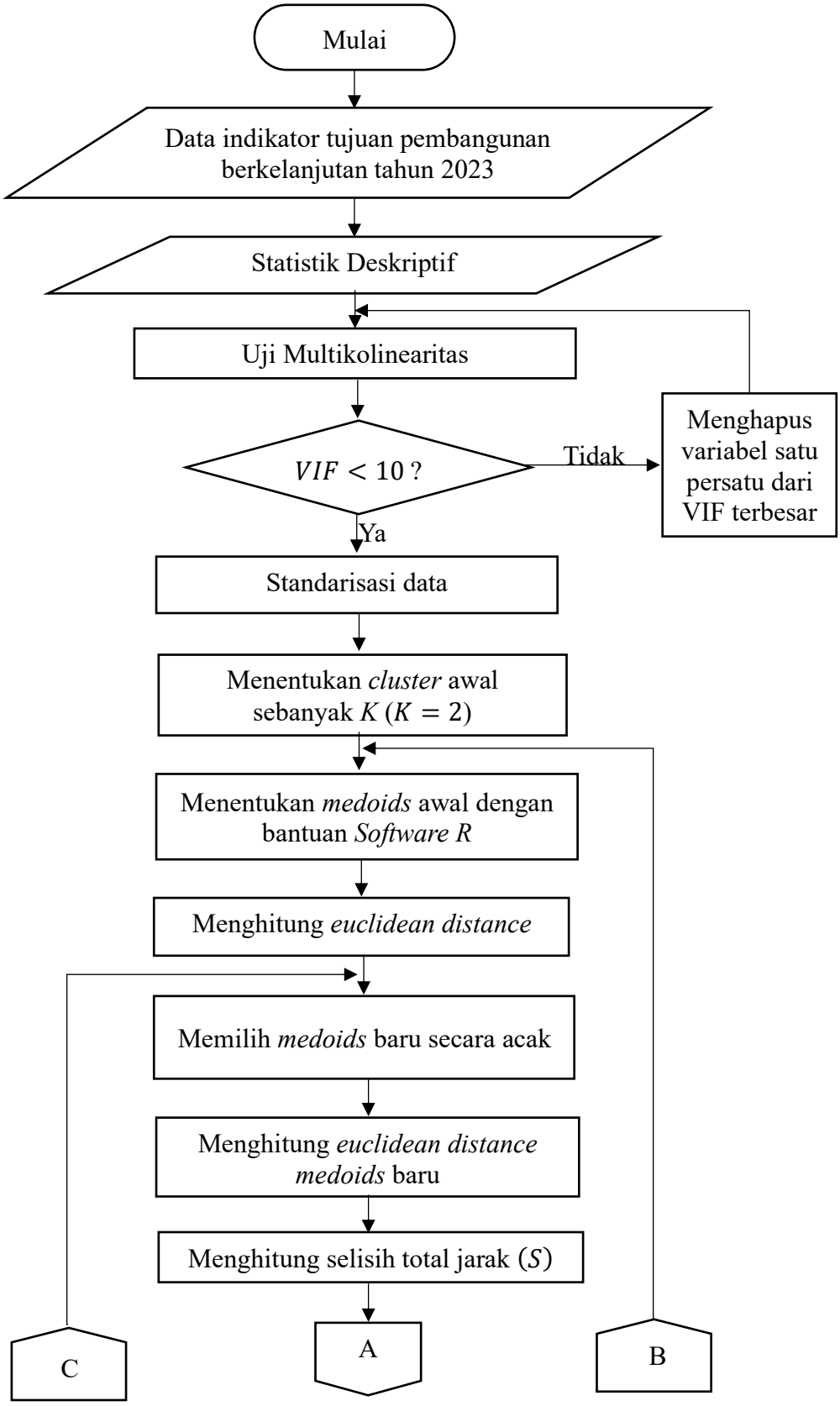
dari *website* Badan Pusat Statistik (BPS), yaitu data indikator tujuan pembangunan berkelanjutan. Variabel penelitian ini terdiri dari 10 indikator yang dibentuk berdasarkan tiga aspek pembangunan berkelanjutan.

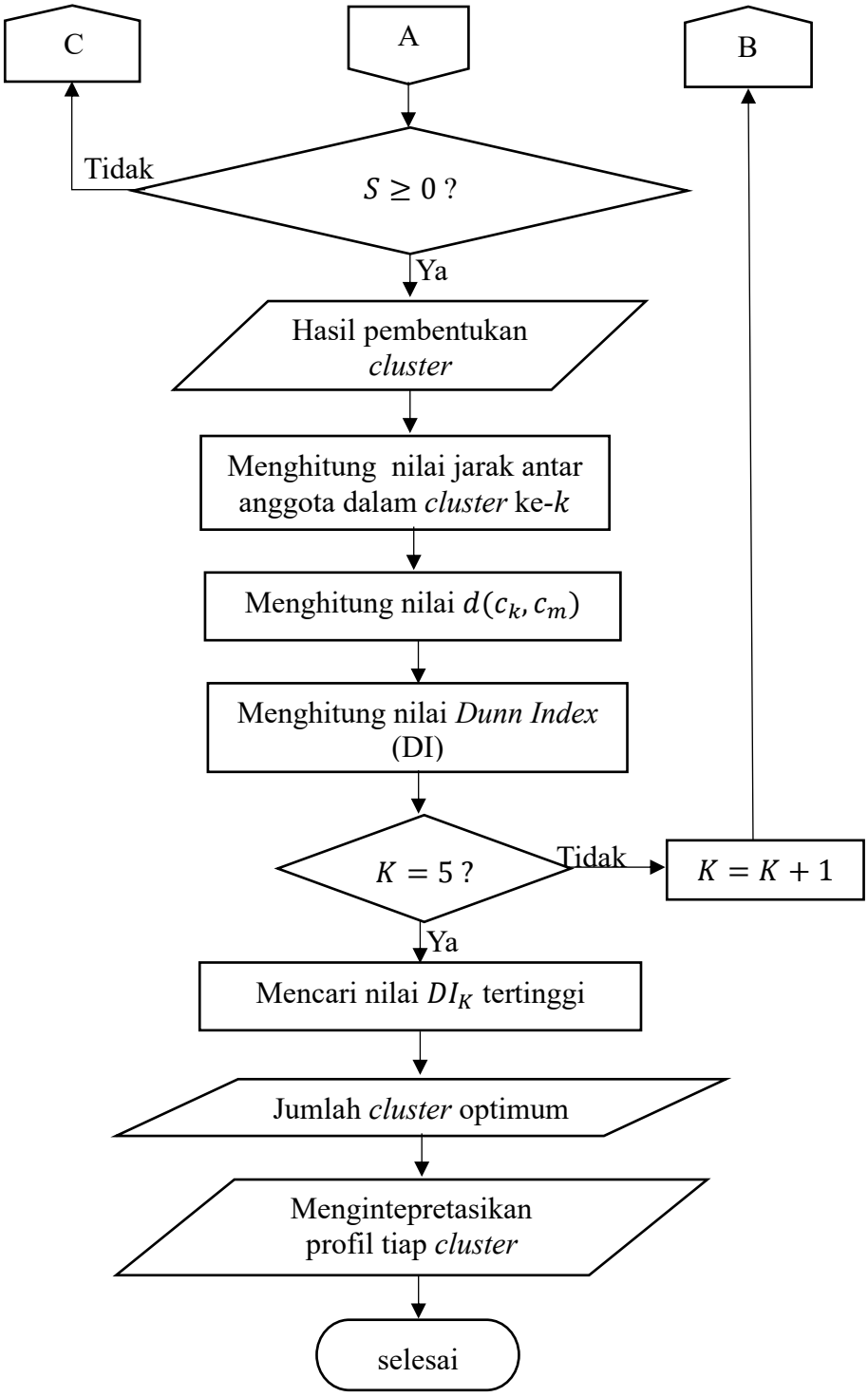
Analisis yang pertama dilakukan adalah analisis statistik deskriptif terhadap data tiap indikator yang digunakan. Tahap berikutnya, melakukan standarisasi data menggunakan *Z-score* dikarenakan data yang digunakan sangat bervariasi dalam satuan pada penelitian ini. Tahap selanjutnya menentukan jumlah *cluster* awal atau kelompok yang akan dibentuk sebanyak 2 buah *cluster*.

Pada algoritma *K-Medoids* untuk tahapan selanjutnya adalah memilih objek amatan dengan bantuan *software R* yang digunakan sebagai *medoids* awal. Kemudian, menghitung jarak *euclidean* dengan objek yang bukan *medoids* dan dimasukkan ke dalam kelompok dengan jarak terdekat terhadap pusat *cluster*. Selanjutnya memilih acak *medoids* baru dan menghitung jarak *euclidean medoids* baru. Setelah itu, menghitung selisih total jarak (S) dan mengulangi kembali seluruh proses hingga iterasi berhenti apabila nilai $S \geq 0$.

Tahap dilanjutkan menghitung diameter *cluster* untuk dua objek dalam satu *cluster.* Kemudian, menghitung nilai $d(c_k, c_m)$ antara *cluster* yang satu dengan *cluster* yang lainnya. Setelah diperoleh nilai diameter *cluster* dan $d(c_k, c_m)$, selanjutnya dilakukan perhitungan nilai *Dunn Index* (DI). Nilai *Dunn index* (DI) dihitung berdasarkan nilai $d(c_k, c_m)$ dan diameter *cluster*. Jika dalam perhitungan sebelumnya dilakukan untuk *cluster* yang bukan berjumlah lima, maka dilakukan perhitungan kembali untuk *cluster* yang berjumlah $K + 1$ dengan tahap menghitung nilai diameter *cluster* hingga menghitung nilai DI. Sedangkan jika perhitungan sebelumnya dilakukan untuk *cluster* yang berjumlah lima, maka tahap dilanjutkan dengan mencari nilai DI terbesar dari hasil beberapa nilai DI sebanyak *cluster* yang telah diperoleh sebelumnya. Setelah jumlah *cluster* optimum yang terbentuk telah diketahui, maka tahap terakhir yang dilakukan adalah menginterpretasikan profil tiap *cluster*.

Langkah-langkah analisis dalam penelitian ini dapat dilihat pada diagram alir disajikan pada Gambar 1.1.


Mulai
Data indikator tujuan pembangunan berkelanjutan tahun 2023
Statistik Deskriptif
Uji Multikolinearitas
$VIF < 10$?
Tidak
Menghapus variabel satu persatu dari VIF terbesar
Ya
Standarisasi data
Menentukan *cluster* awal sebanyak K ($K = 2$)
Menentukan *medoids* awal dengan bantuan *Software R*
Menghitung *euclidean distance*
Memilih *medoids* baru secara acak
Menghitung *euclidean distance medoids* baru
Menghitung selisih total jarak (S)
A
B
C



Gambar 1.1 *Flowchart* Penelitian