OPTIMALISASI JUMLAH *CLUSTER* DENGAN *KRZANOWSKI AND LAI INDEX* BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT DI INDONESIA

SYAKIRA KHANNUR
NIM H1091211028

SKRIPSI

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

OPTIMALISASI JUMLAH *CLUSTER* DENGAN *KRZANOWSKI AND LAI INDEX* BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT DI INDONESIA

SYAKIRA KHANNUR
NIM H1091211028

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	: Optimalisasi Jumlah *Cluster* dengan *Krzanowski and Lai Index* berdasarkan Indikator Kesejahteraan Masyarakat di Indonesia
Nama Mahasiswa	: Syakira Khannur
NIM	: H1091211028
Jurusan/Program Studi	: Matematika / Statistika
Tanggal Lulus	: 30 Januari 2025
SK Pembimbing	: No. 2606/UN22.8/TD.06/2024/Tanggal 2 September 2024
SK Penguji	: No. 59/ UN22.8/TD.06/2025/Tanggal 22 Januari 2025

Dosen Pembimbing

Pembimbing I	Pembimbing II
Dr. Evy Sulistianingsih, M.Sc.	Hendra Perdana, M.Sc.
NIP 198502172008122006	NIP 198810102019031020

Dosen Penguji

Ketua Penguji	Anggota Penguji
Shantika Martha, M.Si.	Nurfitri Imro'ah, M.Si.
NIP 198403082008122003	NIP 198907182019032021

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Dr. Evy Sulistianingsih, M.Sc.	Hendra Perdana, M.Sc.
NIP 198502172008122006	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	: Syakira Khannur
NIM	: H1091211028
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 pernyatan ini untuk digunakan sebagaimana mestinya.

Pontianak, 30 Januari 2025



Syakira Khannur

NIM H1091211028

Optimalisasi Jumlah *Cluster* dengan *Krzanowski and Lai Index* berdasarkan Indikator Kesejahteraan Masyarakat di Indonesia

Abstrak

Kesejahteraan masyarakat merupakan aspek krusial dalam pembangunan suatu negara yang mencerminkan kondisi terpenuhinya kebutuhan dasar penduduk di berbagai aspek, seperti kependudukan, kesehatan, pendidikan, ketenagakerjaan, taraf dan pola konsumsi, perumahan dan lingkungan, serta berbagai indikator sosial lainnya. Tingkat kesejahteraan masyarakat di Indonesia masih menunjukan adanya kesenjangan antar provinsi, terlihat dari beragamnya capaian indikator kesejahteraan di berbagai wilayah. Penelitian ini bertujuan untuk menentukan jumlah *cluster* optimal dalam pengelompokan provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat menggunakan *Krzanowski and Lai Index* (KL *Index*) serta menganalisis karakteristik masing-masing *cluster* yang terbentuk. Metode *K-Medoids* dipilih karena kelebihannya dalam menangani sensitivitas terhadap *outlier* dan *noise*, sedangkan KL *Index* digunakan karena kemampuannya dalam mendeteksi perubahan variasi data. KL *Index* bekerja dengan prinsip bahwa penambahan jumlah *cluster* akan mengurangi variasi dalam *cluster* secara signifikan hingga mencapai jumlah *cluster* yang optimal. Hasil analisis menunjukkan bahwa jumlah *cluster* optimal yang terbentuk adalah 5 *cluster* dengan nilai KL *Index* sebesar 1,8705. *Cluster* 1 dengan tingkat kesejahteraan masyarakat sangat tinggi beranggotakan 1 provinsi yaitu DKI Jakarta. *Cluster* 2 dengan tingkat kesejahteraan masyarakat tinggi beranggotakan 5 provinsi. *Cluster* 3 dengan tingkat kesejahteraan masyarakat sedang beranggotakan 9 provinsi. *Cluster* 4 dengan tingkat kesejahteraan masyarakat rendah beranggotakan 11 provinsi. *Cluster* 5 dengan tingkat kesejahteraan masyarakat sangat rendah beranggotakan 8 provinsi yaitu Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Papua Barat, dan Papua.

Kata kunci: analisis *cluster, outlier, cluster* optimal, *K-Medoids.*

Optimizing The Number Of Clusters Using The Krzanowski and Lai Index Based On Indicators Of Community Welfare In Indonesia

Abstract

Community welfare is a crucial aspect of a country's development that reflects the fulfillment of the basic needs of the population in various aspects, such as population, health, education, employment, consumption levels and patterns, housing, and the environment, as well as various other social indicators. The level of community welfare in Indonesia still shows a gap between provinces, as seen from the diverse achievements of welfare indicators in various regions. This study aims to determine the optimal number of clusters in grouping provinces in Indonesia based on public welfare indicators using the Krzanowski and Lai Index (KL Index) and analyze the characteristics of each cluster formed. The K-Medoids method was chosen for its advantages in handling sensitivity to outliers and noise, while the KL Index was used for its ability to detect changes in data variation. KL Index works on the principle that increasing the number of clusters will significantly reduce the variation within the cluster until it reaches the optimal number of clusters. The results of the analysis show that the optimal number of clusters formed is 5 clusters with a KL Index value of 1.8705. Cluster 1 with a very high level of community welfare consists of 1 province, namely DKI Jakarta. Cluster 2 with a high level of community welfare consists of 5 provinces. Cluster 3 with a medium level of community welfare consists of 9 provinces. Cluster 4 with a low level of community welfare consists of 11 provinces. Cluster 5 with a very low level of community welfare consists of 8 provinces, namely West Nusa Tenggara, East Nusa Tenggara, Central Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, West Papua, and Papua.

Keywords: *cluster analysis, outliers, optimal cluster, K-Medoids.*

PRAKATA

Assalamu'alaikum Warahmatullahi Wabarakatuh

Alhamdulillahirabbil'alamin, puji Syukur kehadirat Allah SWT atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "**Optimalisasi Jumlah *Cluster* dengan *Krzanowski and Lai Index* Berdasarkan Indikator Kesejahteraan Masyarakat di Indonesia**". 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, dukungan, bantuan, 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. Allah SWT yang telah memberikan rahmat, berkah, dan karunia-Nya serta memberikan kelancaran dalam penyusunan tugas akhir ini sehingga dapat diselesaikan dengan baik.
2. Kedua orang tua tercinta, Ibunda Juli Astuti dan Ayahanda Iswanto yang senantiasa memberikan kasih sayang, doa, dukungan moral dan materi, serta motivasi yang tiada hentinya kepada penulis.
3. Ibu Dr. Evy Sulistianingsih, M.Sc. selaku dosen pembimbing pertama yang telah meluangkan waktu untuk memberikan bimbingan, arahan, dan masukan yang sangat berharga dalam penyelesaian tugas akhir ini.
4. Bapak Hendra Perdana, M.Sc. selaku dosen pembimbing kedua yang telah meluangkan waktu untuk memberikan bimbingan, arahan, dan masukan yang sangat berharga dalam penyelesaian tugas akhir ini.
5. Ibu Shantika Martha, M.Si. selaku dosen penguji pertama yang telah memberikan kritik dan saran yang membangun untuk penyempurnaan tugas akhir ini.

6. Ibu Nurfitri Imro'ah, M.Si. selaku dosen penguji kedua yang telah memberikan kritik dan saran yang membangun untuk penyempurnaan tugas akhir ini.
7. Ibu Mariatul Kiftiah, M.Sc. dan Ibu Wirda Andani, M.Si. selaku dosen pembimbing akademik yang telah membimbing dan mengarahkan penulis selama masa perkuliahan.
8. Keluarga besar yang penulis sayangi, terutama Alm Asmuni Hasan dan Rohani selaku kakek dan nenek penulis yang selalu memberikan dukungan dan semangat. Penulis juga berterima kasih kepada saudari-saudari penulis yaitu Nisrina, Shafa, dan Hanna, yang senantiasa memberikan motivasi dan perhatian dalam setiap langkah penulis.
9. Rekan-rekan seperjuangan yang telah membersamai penulis dalam menimba ilmu di Program Studi Statistika yaitu Setio, Dea, dan Nirwani.
10. Sahabat sejati penulis, Najwa Sari Fauziah yang telah memberikan dukungan, semangat, serta menjadi tempat ternyaman bagi penulis untuk berkeluh-kesah. Penulis juga mengucapkan terima kasih kepada sahabat-sahabat lainnya, yaitu Naufal, Bagus, Azhar, Windi, dan Julian, atas segala dukungan dan kebersamaan yang telah terjalin.
11. Seluruh mahasiswa Statistika Universitas Tanjungpura, khususnya "Expos'21", yang namanya tidak dapat penulis sebutkan satu per satu yang telah menjadi bagian dari perjalanan studi penulis.
12. Seluruh pihak yang tidak dapat penulis sebutkan satu per satu yang telah membantu dalam penyelesaian tugas akhir ini.

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

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Pontianak, 30 Januari 2025

Syakira Khannur

DAFTAR ISI

Abstrak iv
Abstract v
PRAKATA vi
DAFTAR ISI viii
DAFTAR TABEL x
DAFTAR GAMBAR xi
DAFTAR LAMPIRAN xii
DAFTAR SIMBOL xiii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 4
1.4 Batasan Masalah 4
1.5 Tinjauan Pustaka 5
1.6 Metodologi Penelitian 6
BAB II LANDASAN TEORI 11
2.1 Indikator Kesejahteraan Masyarakat 11
2.2 Analisis Multivariat 11
2.3 Analisis *Cluster* 12
2.4 Metode pembentukan *Cluster* 13
2.4.1 Metode Pengelompokan Hierarki 13
2.4.2 Metode Pengelompokan Non-Hierarki 14
BAB III OPTIMALISASI JUMLAH *CLUSTER* 16
3.1 Asumsi Analisis *Cluster* 16
3.2 Tahapan Analisis *Cluster* 17
3.3 Algoritma *K-Medoids* 19
3.4 *Krzanowski and Lai Index* 21
BAB IV HASIL DAN PEMBAHASAN 23
4.1 Deskripsi Data 23
4.2 Asumsi Analisis *Cluster* 25
4.3 Standarisasi Data 26
4.4 Pengelompokan dengan Algoritma *K-Medoids* 26

4.4.1 Penentuan *Medoids* Awal 26
4.4.2 Penentuan *Medoids* Baru 28
4.4.3 Menghitung Selisih Total Jarak Dalam *Cluster* 29
4.5 Penentuan Jumlah *Cluster* Optimal dengan *Krzanowski and Lai Index* 29
4.6 Interpretasi Profil *Cluster* 32
BAB V KESIMPULAN 39
DAFTAR PUSTAKA 41
LAMPIRAN 45

DAFTAR TABEL

Tabel 4.1 Variabel Penelitian 23
Tabel 4.2 Statistik Deskriptif 24
Tabel 4.3 Nilai *Variance Inflation Factor* 25
Tabel 4.4 *Medoids* Awal Untuk $k = 3$ 27
Tabel 4.5 Jarak Aceh dan Jambi ke setiap *Medoids* 28
Tabel 4.6 *Medoids* Baru Untuk $k = 3$ 28
Tabel 4.7 Pengelompokan Provinsi di Indonesia dengan Tiga *Cluster* 29
Tabel 4.8 Optimalisasi Jumlah *Cluster* dengan *Krzanowski and Lai Index* 31
Tabel 4.9 Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat 32
Tabel 4.10 Rata-Rata Variabel Indikator Kesejahteraan Masyarakat 33

DAFTAR GAMBAR

Gambar 1.1 *Flowchart* Optimalisasi Jumlah *Cluster* 9
Gambar 4.1 Plot Nilai *Krzanowski and Lai Index* untuk setiap *k* 31
Gambar 4.2 Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat 35

DAFTAR LAMPIRAN

Lampiran I. Data Indikator Kesejahteraan Masyarakat Tiap Provinsi di Indonesia Tahun 2023..... 45
Lampiran II. Data Hasil Standarisasi 46
Lampiran III. Jarak Objek ke *Medoids* pada saat $k = 3$..... 48
Lampiran IV. Jarak Objek ke *Medoids* pada saat $k = 5$ 49
Lampiran V. Nilai W(k), DIFF(k), dan KL(k) untuk setiap k 50

DAFTAR SIMBOL

I	:	Banyak objek penelitian
i	:	Objek ke-i dengan i = 1, 2, 3, ..., I
K	:	Banyak *cluster*
k	:	*Cluster* ke-k dengan k = 1, 2, 3, …, K
J	:	Banyak variabel penelitian
j	:	Variabel ke-j dengan j = 1, 2, 3, …, J
$Z_{i;j}$	:	Standarisasi data untuk objek ke-i pada variabel ke-j
x_i		Nilai objek ke-i
$x_{i;j}$	:	Nilai dari objek ke-i pada variabel ke-j
$x_{f;j}$	:	Nilai objek ke-f pada variabel ke-j
μ_j	:	Rata-rata variabel ke-j
σ_j	:	Standar deviasi variabel ke-j
$r_{j;h}$	:	Koefisien korelasi antara variabel ke-j dan ke-h
$a_{j;h}$	:	Koefisien korelasi parsial antara variabel ke-j dan ke-h
R_j^2	:	Koefisien determinasi variabel ke-j terhadap variabel independen lainnya
$d(x_i; x_f)$	:	Jarak *euclidean* antara objek ke-i dan objek ke-f
$d(x_i; M_m)$	:	Jarak *euclidean* antara objek ke-i dan *medoids* ke-m
$M_{m;j}$	:	*Medoids* ke-m pada variabel ke-j
C_m	:	*Cluster* ke-m yang berisi kumpulan objek yang terkelompok dengan *medoids* ke-m.
$KL\ (k)$	:	Nilai *Krzanowski and Lai Index* untuk setiap *cluster* ke-k
$DIFF\ (k)$	:	Fungsi yang mengukur perbedaan antara jumlah kuadrat dalam *cluster* untuk $(k-1)$ dan k
$W(k)$	:	Jumlah kuadrat dalam *cluster* k

BAB I
PENDAHULUAN

1.1 Latar Belakang

Rakyat adalah komponen utama dalam pembentukan dan keberlangsungan suatu negara. Suatu negara tidak dapat terbentuk tanpa adanya rakyat (Rombot, Senewe dan Paseki, 2023). Peran rakyat sangat signifikan dalam membentuk, menjaga, dan mengembangkan negara. Suatu negara harus menjamin kesejahteraan rakyatnya karena kesejahteraan rakyat berpengaruh secara signifikan pada sistem politik, ekonomi, dan sebagainya (Susanto, Puspitaningrum dan Neneng, 2021). Kesejahteraan rakyat adalah suatu kondisi dimana rakyat dapat dengan mudah memperoleh kebutuhan dasar yang memadai, seperti kebutuhan pangan, kesehatan, keamanan, pendidikan, dan pendapatan yang dapat memberikan kehidupan yang layak. Kesejahteraan masyarakat pada dasarnya merupakan suatu kondisi sosial yang memungkinkan bagi setiap warga negara untuk mampu memenuhi kebutuhan hidup jasmani, mental, dan sosial sesuai dengan fitrah dan martabat sebagai manusia agar dapat mengatasi setiap permasalahan sosial yang dihadapi individu, kelompok, dan masyarakatnya agar dapat berkembang menjadi lebih baik (Rahman, 2018).

Mengingat betapa pentingnya peran rakyat bagi suatu negara, memunculkan beberapa teori yang memfokuskan perhatian kepada kesejahteraan rakyat, salah satunya teori kedaulatan rakyat. Hakikat dari kedaulatan adalah suatu konsep tentang kekuasaan tertinggi disuatu negara dan tidak dapat diintervensi atau bahkan dikurangi oleh negara lain (Qoroni dan Winarwati, 2021). Teori kedaulatan rakyat menjelaskan bahwa kedaulatan negara berada ditangan rakyat, artinya pemegang kekuasaan tertinggi di negara yang menganut adalah rakyat. Indonesia adalah salah satu negara yang menganut teori kedaulatan rakyat (Kaingge, 2017). Konsep ini tercermin dalam Pasal 1 ayat (2) UUD 1945 yang menyatakan bahwa “Kedaulatan berada ditangan rakyat dan dilaksanakan menurut Undang-Undang Dasar”. Selain

itu, sila ke-4 Pancasila yang menegaskan bahwa “Kerakyatan yang dipimpin oleh hikmah kebijaksanaan dalam permusyawaratan perwakilan”.

Tingkat kesejahteraan masyarakat dapat dilihat melalui berbagai indikator yang meliputi berbagai aspek penting kehidupan. Berdasarkan publikasi Badan Pusat Statistik yang berjudul Indikator Kesejahteraan Rakyat 2023, terdapat beberapa aspek dalam mengukur tingkat kesejahteraan masyarakat di Indonesia. Beberapa aspek yang dimaksud meliputi aspek kependudukan, kesehatan dan gizi, pendidikan, ketenagakerjaan, taraf dan pola konsumsi, perumahan dan lingkungan, kemiskinan, dan sosial lainnya yang menjadi acuan dalam upaya peningkatan kualitas hidup (Badan Pusat Statistik, 2023).

Pemerintah telah melakukan berbagai upaya dalam bidang pembangunan agar negara Indonesia dapat mewujudkan kesejahteraan bagi seluruh rakyatnya melalui berbagai bidang pembangunan tersebut (Basofi dan Santoso, 2017). Saat ini telah banyak program pemerintah yang bertujuan untuk meningkatkan kesejahteraan rakyat seperti Program Keluarga Berencana, Perlindungan Sosial seperti Kartu Indonesia Pintar, Program Keluarga Harapan dan program-program lainnya. Namun hingga saat ini perwujudan dari kesejahteraan rakyat tersebut belum maksimal. Hal tersebut dikarenakan ketidakmerataan pembangunan sehingga menimbulkan ketimpangan pada setiap daerah. Menurut Badan Pusat Statistik (2023), Indeks Pembangunan Manusia (IPM) DKI Jakarta mencapai 83,55 sementara Papua hanya 63,01. Disparitas sosial ini menunjukkan adanya kesenjangan dalam hal kesejahteraan antar provinsi di Indonesia.

Kesejahteraan rakyat juga merupakan bagian dari usaha pencapaian beberapa target *Sustainable Development Goals* (SDGs) di Indonesia (Solomon, 2023). Beberapa target SDGs yang relevan dengan indikator kesejahteraan rakyat antara lain SDG 1 (Tanpa Kemiskinan), SDG 3 (Kehidupan Sehat dan Sejahtera), SDG 4 (Pendidikan Berkualitas), dan SDG 8 (Pekerjaan Layak dan Pertumbuhan Ekonomi). Oleh karena itu, upaya peningkatan kesejahteraan rakyat tidak hanya penting bagi pembangunan nasional, tetapi juga berkontribusi pada pencapaian agenda global.

Untuk memahami dan mengatasi kesenjangan kesejahteraan antar provinsi di Indonesia, analisis *cluster* merupakan metode yang relevan. Analisis *cluster* memungkinkan pengelompokan provinsi berdasarkan karakteristik yang serupa, hal ini dapat membantu pemerintah dalam perumusan kebijakan yang lebih tepat. Pada penelitian ini dilakukan pengelompokan provinsi di Indonesia berdasarkan indikator kesejahteraan rakyat. Pengelompok objek menggunakan metode *K-Medoids* dengan jarak *euclidean*. *K-Medoids* merupakan pengembangan dari algoritma *K-Means* yang lebih *robust* terhadap *noise* dan *outlier.* Analisis terhadap data indikator kesejahteraan masyarakat di Indonesia menunjukkan adanya variasi yang sangat besar antar provinsi, bahkan beberapa provinsi memiliki nilai yang ekstrem (*outlier*) pada indikator tertentu. Sebagai contoh, terdapat kesenjangan yang signifikan pada kepadatan penduduk antar provinsi, dimana DKI Jakarta memiliki kepadatan penduduk yang jauh lebih tinggi dibandingkan provinsi lainnya. Demikian pula dengan persentase penduduk miskin yang menunjukkan variasi yang cukup besar antar provinsi. Untuk memastikan bahwa analisis *cluster* yang telah dilakukan efektif maka diperlukan jumlah *cluster* optimal. Terdapat banyak metode dalam menentukan jumlah *cluster* optimal seperti *Elbow*, *Silhouette*, *Calinski Harabasz, Pseudo-F Statistic*, *Dunn Index,* dan *Davies Bouldin Index* (Shalsadilla *et al.*, 2023).

Penelitian ini menggunakan *Krzanowski and Lai Index* atau KL *Index* sebagai acuan dalam menentukan jumlah *cluster* optimal. Pemilihan indeks ini didasarkan pada kemampuannya dalam menangani data multidimensi dan kestabilannya dalam menentukan jumlah *cluster* optimal (Krzanowski dan Lai, 1988). Indeks ini diperkenalkan pada tahun 1988 oleh W. J. Krzanowski dan Y. T. Lai untuk menemukan titik optimal dari jumlah *cluster*, dengan cara menganalisis perubahan variasi antar-*cluster* dan intra-*cluster* seiring bertambahnya jumlah *cluster.*

Hasil dari penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas terkait pengelompokan provinsi berdasarkan tingkat kesejahteraan rakyat. Dengan mempertimbangkan karakteristik dan kebutuhan unik masing-masing kelompok provinsi, informasi ini dapat dimanfaatkan pemerintah dalam membuat kebijakan pembangunan yang lebih tepat sasaran. Dengan demikian, kesejahteraan

rakyat dapat ditingkatkan dengan lebih efisien dan efektif, mengurangi kesenjangan di setiap provinsi, dan mampu mendukung pencapaian tujuan pembangunan nasional dan SDGs.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah dalam penelitian ini, yaitu:

1. Bagaimana penentuan jumlah *cluster* optimal pada pengelompokan provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat tahun 2023 menggunakan *Krzanowski and Lai Index*?
2. Bagaimana karakteristik masing-masing *cluster* yang terbentuk?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disusun, adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Penelitian ini bertujuan untuk menentukan jumlah *cluster* optimal pada pengelompokan provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat tahun 2023 menggunakan *Krzanowski and Lai Index.*
2. Penelitian ini juga bertujuan untuk menganalisis karakteristik masing-masing *cluster* yang terbentuk**.**

1.4 Batasan Masalah

Batasan masalah digunakan sebagai acuan agar penelitian lebih terarah dan tidak terjadi penyimpangan. Penelitian ini hanya membahas data indikator kesejahteraan masyarakat tahun 2023 yang bersumber dari Badan Pusat Statistik (BPS) Indonesia. Data yang digunakan yaitu kepadatan penduduk per $KM^2(X_1)$, laju pertumbuhan penduduk (X_2), umur harapan hidup (X_3), rata-rata lama sekolah (X_4), tingkat pengangguran terbuka (X_5),tingkat partisipasi angkatan kerja (X_6), pengeluaran perkapita disesuaikan (X_7), indeks gini (X_8), kepemilikan rumah sendiri (X_9), rumah tangga dengan akses air layak (X_{10}), dan persentase penduduk

miskin (X_{11}). Pengelompokan data dilakukan pada 34 provinsi di Indonesia menggunakan analisis *cluster* non-hierarki dengan metode *K-Medoids*. Penentuan jumlah *cluster* optimal menggunakan metode *Krzanowski and Lai Index* (KL *Index*), dengan jumlah *cluster* optimal yang akan diuji berkisar antara 2 hingga 32 *cluster*.

1.5 Tinjauan Pustaka

Novianti, Setyorini dan Rafflesia (2017) melakukan penelitian yang bertujuan untuk mengelompokkan episenter gempa di Provinsi Bengkulu menggunakan data gempa bumi tektonik dari Januari 1970 hingga Desember 2015 yang diperoleh dari stasiun BMKG Kepahiang Bengkulu. Metode *K-Means* dengan jarak *euclidean* digunakan untuk analisis ini, dengan variabel lintang, bujur, dan magnitudo. Jumlah *cluster* optimal ditentukan dengan *Krzanowski and Lai Index*, yang menunjukkan 7 *cluster*.

Rudianto dan Wijayanto (2023) melakukan penelitian yang bertujuan untuk pengelompokan (clustering) 34 provinsi di Indonesia berdasarkan aspek yang terdapat dalam IDI Tahun 2021. Metode *clustering* yang digunakan dalam penelitian ini yaitu *K-Means* dan *K-Medoids* dengan optimalisasi jumlah *cluster* menggunakan *Dunn Index, Silhouette, Calinski-Harabasz, dan Davies-Bouldin*. Hasil penelitian menunjukkan bahwa algoritma *K-Medoids* lebih baik karena rasio Sw/Sb lebih kecil dibandingkan *K-Means* dengan jumlah cluster optimal sebanyak 5 *cluster.*

Siagian, Sirait, dan Halim (2022) melakukan penelitian yang bertujuan untuk menerapkan metode *K-Means* dan *K-Medoids* untuk mengelompokkan pelanggan berdasarkan data transaksi *e-commerce*. Hasil penelitian menunjukkan bahwa kedua algoritma menunjukkan hasil *cluster* optimal yang sama yaitu k = 3. Hasil pengujian juga menunjukkan bahwa *K-Medoids* memiliki performa terbaik dengan nilai ratio sebesar 0,337 dibandingkan dengan *K-Means* 0,338.

Nabilah, Perdana dan Sulistianingsih (2024) melakukan penelitian yang bertujuan untuk mengelompokkan provinsi di Indonesia berdasarkan indikator kesejahteraan masyarakat dengan metode *K-Means* ++. Data yang digunakan

merupakan 10 indikator kesejahteraan masyarakat tahun 2022 dan menggunakan *silhouette coefficient* untuk menganalisis jumlah *cluster* optimal. Hasil penelitian menunjukkan bahwa jumlah *cluster* optimal yang terbentuk adalah 3 *cluster. Cluster* 1 yang terdiri atas 27 provinsi merupakan *cluster* dengan tingkat kesejahteraan masyarakat sedang. *Cluster* 2 yang terdiri atas 3 provinsi merupakan *cluster* dengan tingkat kesejahteraan masyarakat rendah, dan *cluster* 3 yang terdiri atas 4 provinsi merupakan *cluster* dengan tingkat kesejahteraan masyarakat tinggi.

Belinda, HG ,dan Yozza (2019) melakukan penelitian yang bertujuan untuk mengelompokan provinsi di Indonesia berdasarkan indikator kesejahteraan menggunakan metode *cluster ensemble* dengan algoritma *Cluster Ensemble Based Mixed Data Clustering* (CEBMDC). Data numerik dikelompokkan dengan metode *Ward*, sedangkan data numerik dikelompokkan dengan metode *Robust Clustering Using Links* (ROCK). Hasil penelitian tersebut menunjukkan bahwa jumlah *cluster* optimal yang terbentuk adalah 4 *cluster. Cluster* 1 terdiri atas 1 provinsi yaitu Gorontalo, *cluster* 2 terdiri atas 19 provinsi, *cluster* 3 terdiri atas 10 provinsi, dan *cluster* 4 terdiri atas 4 provinsi. *Cluster* 2 yang memiliki anggota terbanyak memiliki karakteristik yang lebih unggul kesejahteraan rakyatnya dibandingkan *cluster* lainnya, terutama dalam bidang pendidikan, ekonomi, dan kesehatan.

1.6 Metodologi Penelitian

Langkah pertama dalam penelitian ini adalah pengumpulan data. Data yang digunakan merupakan indikator kesejahteran rakyat untuk setiap provinsi di Indonesia tahun 2023 yang bersumber dari Badan Pusat Statistik. Terdapat beberapa aspek yang ditinjau untuk menentukan kesejahteraan rakyat seperti aspek kependudukan, aspek kesehatan, aspek pendidikan, aspek ketenagakerjaan, aspek taraf dan pola konsumsi, aspek perumahan dan lingkungan, aspek kemiskinan, dan aspek sosial lainnya.

Proses analisis dilakukan dengan bantuan *software RStudio*. Langkah pertama dalam proses analisis adalah melakukan *preprocessing data* dengan tujuan untuk memastikan tidak terdapat *missing data*. *Missing data* atau data yang hilang dapat menyebabkan bias dalam hasil analisis dan mengurangi akurasi model. Pengecekan

missing data dilakukan dengan menggunakan fungsi yang tersedia di *RStudio* untuk mengidentifikasi sel-sel kosong atau nilai NA (*Not Available*) dalam dataset.

Langkah selanjutnya adalah analisis statistik deskriptif untuk setiap variabel. Statistik deskriptif bertujuan untuk menyajikan dan merangkum data secara sistematis serta memberikan pemahaman yang lebih mendalam mengenai karakteristik atau sifat dasar dari variabel yang diamati. Melalui analisis ini, diharapkan dapat membentuk gambaran yang jelas tentang dataset yang tersedia serta memberikan landasan untuk penelitian lebih lanjut atau pengambilan keputusan.

Setelah melakukan analisis statistik deskriptif, langkah selanjutnya adalah melakukan uji multikolinearitas untuk memastikan bahwa variabel yang digunakan tidak memiliki korelasi yang tinggi antar satu sama lain. Dalam analisis *cluster*, korelasi yang tinggi antar variabel dapat menyebabkan beberapa masalah, seperti redundansi informasi, di mana variabel yang sangat berkorelasi membawa informasi serupa sehingga algoritma *clustering* memberikan bobot berlebih pada pola yang berulang. Hal ini dapat mendistorsi hasil *clustering* karena pola dominan dari variabel berkorelasi tinggi mendikte pembentukan *cluster*, sehingga hasil akhirnya kurang mencerminkan variasi yang ada di data.

Langkah selanjutnya adalah melakukan standarisasi data, yaitu proses mengubah nilai suatu kumpulan data ke format standar tertentu. Tujuannya untuk memastikan setiap variabel memiliki skala yang sama, sehingga tidak ada variabel yang mendominasi dalam proses pembentukan *cluster*. Proses ini melibatkan perubahan setiap nilai dalam kumpulan data agar memiliki mean nol dan standar deviasi satu, atau menggunakan skala lain yang dianggap lebih mudah untuk diinterpretasikan atau dibandingkan.

Kemudian mengelompokan objek menggunakan analisis non-hierarki dengan algoritma *K-Medoids*. Dalam penelitian ini, objek yang dimaksud adalah setiap provinsi di Indonesia. Langkah pertama dalam pengelompokan objek dengan algoritma *K-Medoids* adalah menentukan jumlah *cluster* yang akan dianalisis. Jumlah *cluster* yang akan diuji dalam penelitian ini adalah 2 hingga 32. Pemilihan rentang 2 hingga 32 didasarkan pada ketentuan KL *Index* yang memerlukan nilai

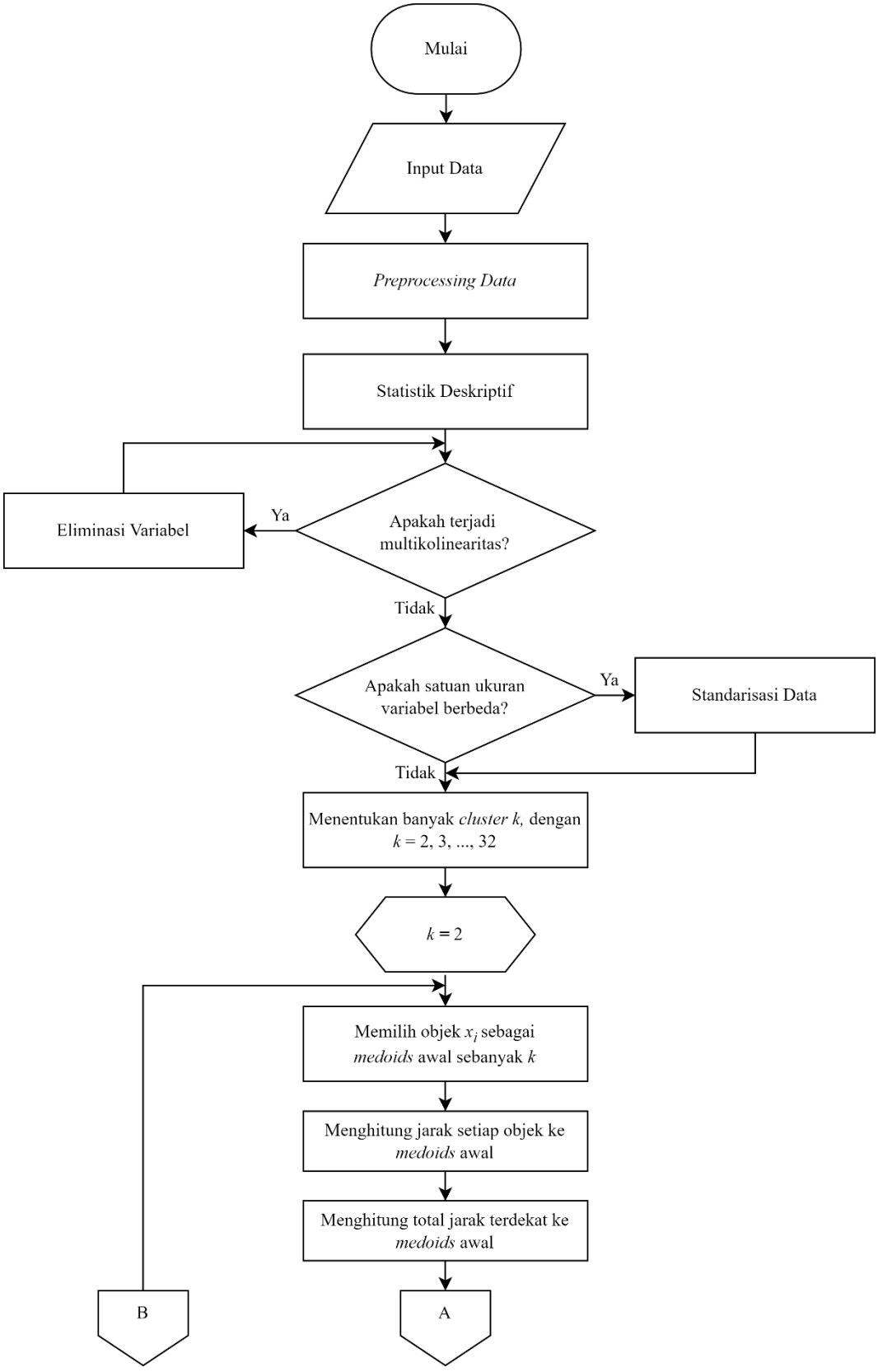
(*k-1)* dan *(k+1)* untuk setiap perhitungan nilai k dan disesuaikan dengan jumlah objek yang dianalisis.

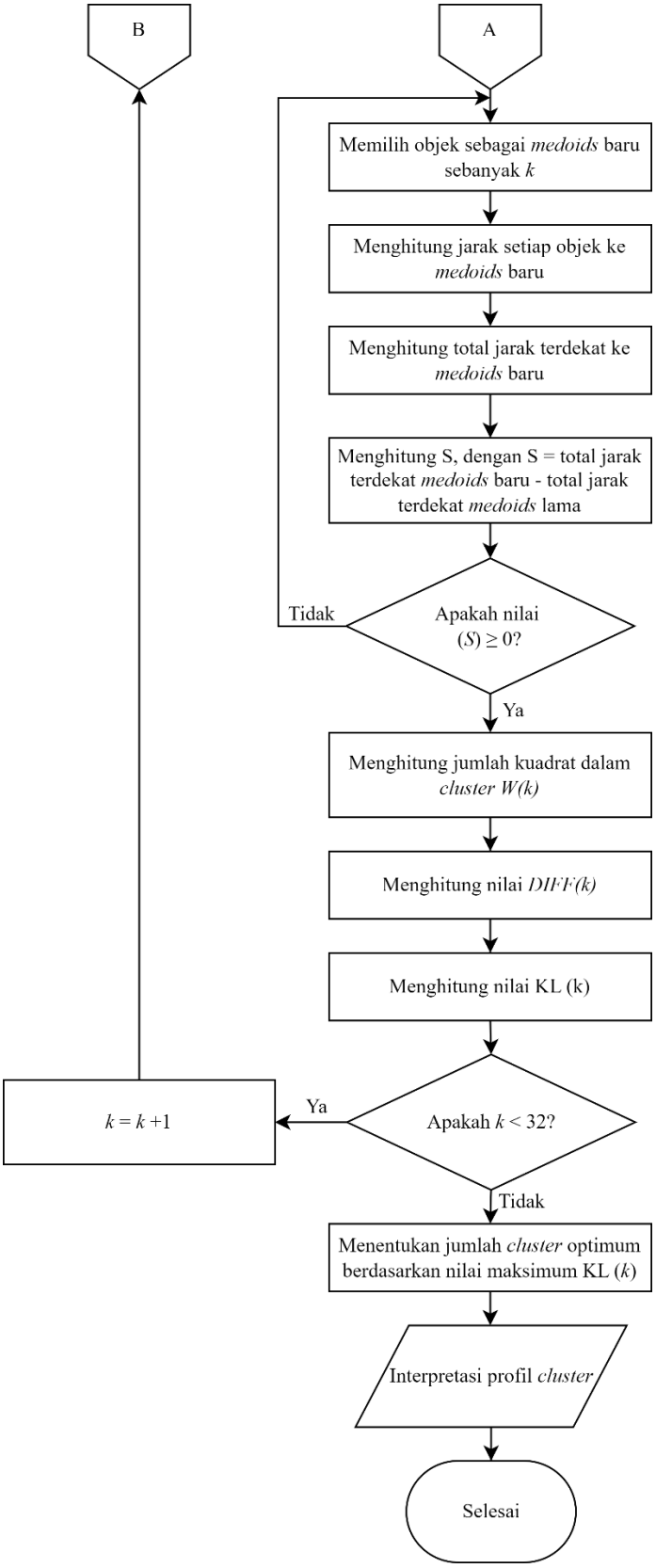
Langkah selanjutnya adalah memilih objek x_i sebagai *medoids* awal sebanyak *k,* kemudian menghitung jarak *euclidean* antara objek dengan *medoids* yang telah dipilih. Berdasarkan perhitungan jarak tersebut, setiap objek dimasukkan ke dalam *cluster* yang memiliki jarak terdekat dengan *medoids*. Kemudian menghitung total jarak terdekat setiap objek.

Langkah berikutnya adalah memilih *medoids* baru sebanyak *k*, kemudian menghitung jarak *euclidean* dan total jarak terdekat setiap objek ke *medoids* baru. Setelah itu, menghitung selisih total jarak terdekat *medoids* awal dan *medoids* baru (S). Jika nilai $S \geq 0$, maka proses iterasi berhenti.

Kemudian menghitung nilai *W(k)* dan *DIFF(k)* dengan *k* adalah jumlah *cluster* dengan rentang 2 hingga 32. Setelah itu melakukan analisis untuk menentukan jumlah *cluster* optimal berdasarkan *Krzanowski and Lai Index*. Langkah selanjutnya dalam penelitian ini adalah melakukan interpretasi untuk setiap *cluster* yang terbentuk dengan mengidentifikasi karakteristik dari masing-masing *cluster.*

Penelitian ini dilakukan dengan menggambarkan alur penyelesaian yang jelas dengan langkah-langkah yang diperlukan melalui diagram alir yang digambarkan sebagai berikut.


Mulai
Input Data
Preprocessing Data
Statistik Deskriptif
Apakah terjadi multikolinearitas?
Ya
Eliminasi Variabel
Tidak
Apakah satuan ukuran variabel berbeda?
Ya
Standarisasi Data
Tidak
Menentukan banyak *cluster k,* dengan $k = 2, 3, ..., 32$
$k = 2$
Memilih objek x_i sebagai *medoids* awal sebanyak k
Menghitung jarak setiap objek ke *medoids* awal
Menghitung total jarak terdekat ke *medoids* awal
B
A



Gambar 1.1 *Flowchart* Optimalisasi Jumlah *Cluster*