METODE *SILHOUETTE COEFFICIENT* UNTUK PENENTUAN JUMLAH *CLUSTER* OPTIMUM PADA PENGELOMPOKAN PROVINSI DI INDONESIA
(Studi Kasus: Indikator Pendidikan Tingkat SMA/Sederajat)

PUTRI RAHMADANTI
NIM H1091211051

SKRIPSI



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

METODE *SILHOUETTE COEFFICIENT* UNTUK PENENTUAN JUMLAH *CLUSTER* OPTIMUM PADA PENGELOMPOKAN PROVINSI DI INDONESIA
(Studi Kasus: Indikator Pendidikan Tingkat SMA/Sederajat)

PUTRI RAHMADANTI
NIM H1091211051

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	: Metode *Silhouette Coefficient* Untuk Penentuan Jumlah *Cluster* Optimum Pada Pengelompokan Provinsi Di Indonesia
Nama Mahasiswa	: Putri Rahmadanti
NIM	: H1091211051
Jurusan/Program Studi	: Matematika/ Statistika
Tanggal Lulus	: 23 Januari 2025
SK Pembimbing	: No.2626/UN22.8/TD.06/2024/2 September 2024
SK Penguji	: No.5563/UN22.8/TD.06/2024/31 Desember 2024

Dosen Pembimbing

Pembimbing I	Pembimbing II
Shantika Martha, M.Si.	Neva Satyahadewi, M.Sc., CRA., CRP., CRMP.
NIP 198403082008122003	NIP 198212042005012001

Dosen Penguji

Ketua Penguji	Anggota Penguji
Nurfitri Imro'ah, M.Si.	Hendra Perdana, M.Sc.
NIP 198907182019032021	NIP 198810102019031020

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Shantika Martha, M.Si.	Neva Satyahadewi, M.Sc., CRA., CRP., CRMP.
NIP 198403082008122003	NIP 198212042005012001

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

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

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama	: Putri Rahmadanti
NIM	: H1091211051
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, 23 Januari 2025



Putri Rahmadanti
H1091211051

Metode *Silhouette Coefficient* Untuk Penentuan Jumlah *Cluster* Optimum Pada Pengelompokan Provinsi Di Indonesia (Studi Kasus: Indikator Pendidikan Tingkat SMA/Sederajat)

Abstrak

Pendidikan merupakan aspek yang berkaitan dengan kualitas sumber daya manusia yang memiliki peran penting dalam upaya peningkatan pembangunan nasional. Pada dasarnya pemerataan pendidikan sangat krusial untuk meningkatkan pendidikan di Indonesia. Indonesia yang terdiri dari banyak provinsi dengan beragam kondisi menjadikan perlu dilakukan pengelompokan terhadap pendidikan pada tiap daerah. Analisis *cluster* adalah teknik analisis statistik multivariat yang melakukan pengelompokan berdasarkan atas kesamaan karakteristik antar tiap objeknya. Tujuan penelitian ini adalah menentukan jumlah *cluster* optimum menggunakan metode *Silhouette Coefficient* pada pengelompokan provinsi di Indonesia berdasarkan indikator pendidikan tingkat SMA/sederajat. Data yang digunakan adalah data indikator pendidikan tingkat SMA/sederajat yang terdiri dari 11 variabel dari 34 provinsi di Indonesia. Tahapan penelitian ini meliputi pengelompokan dengan menggunakan metode *Ward* dan dilanjutkan dengan proses penentuan jumlah *cluster* optimum menggunakan metode *Silhouette Coefficient*. Berdasarkan analisis yang telah dilakukan pada pengelompokan tingkat pendidikan di Indonesia, diperoleh *cluster* optimum berdasarkan indikator pendidikan tingkat SMA/sederajat, yaitu berjumlah 3 *cluster* dengan nilai *Silhouette Coefficient* sebesar 0,248. *Cluster* ke-1 merupakan *cluster* dengan tingkat pendidikan tinggi terdiri dari 8 provinsi yang didominasi dengan provinsi yang berada di Pulau Jawa. Kemudian *cluster* ke-2 sebagai *cluster* dengan tingkat pendidikan sedang terdiri dari 25 provinsi yang merupakan *cluster* terbesar yang mewakili sebagian besar wilayah di Indonesia. Serta *cluster* ke-3 sebagai *cluster* dengan tingkat pendidikan rendah hanya terdiri dari Provinsi Papua yang menunjukkan kesenjangan pendidikan yang signifikan.

Kata kunci: Jarak *Euclidean*, Metode *Ward*, *Z-Score*.

Silhouette Coefficient Method for Determining the Optimum Number of Clusters in Grouping Provinces in Indonesia (Case Study: High School Education Indicators)

Abstract

Education is an aspect related to the quality of human resources that has an important role in efforts to improve national development. Basically, equal distribution of education is crucial to improve education in Indonesia. Indonesia, which consists of many provinces with various conditions, makes it necessary to group education in each region. Cluster analysis is a multivariate statistical analysis technique that performs grouping based on the similarity of characteristics between each object. The purpose of this study is to determine the optimal number of clusters using the Silhouette Coefficient method in the grouping of provinces in Indonesia based on high school/equivalent education indicators. The data used is data on high school/equivalent education indicators consisting of 11 variables from 34 provinces in Indonesia. The stages of this research include grouping using the Ward method and followed by the process of determining the optimal number of clusters using the Silhouette Coefficient method. Based on the analysis that has been carried out on the grouping of education levels in Indonesia, the optimum cluster was obtained based on the education indicators at the high school level/equivalent, which amounted to 3 clusters with a Silhouette Coefficient value of 0.248. The 1st cluster is a cluster with a higher education level consisting of 8 provinces dominated by provinces on the island of Java. Then the 2nd cluster as a cluster with a medium level of education consists of 25 provinces which is the largest cluster representing most of the regions in Indonesia. And the 3rd cluster as a cluster with a low level of education only consists of Papua Province which shows a significant educational gap.

Keywords: *Euclidean Distance, Ward Method, Z-Score*.

PRAKATA

Dengan mengucapkan puji dan syukur kepada Allah *Subhanahu wata'ala* yang mana atas berkah, karunia, serta kasih sayang-Nya penulis dapat menyelesaikan skripsi ini dengan baik. Adapun judul skripsi yang diajukan yaitu "**Metode *Silhouette Coefficient* Untuk Penentuan Jumlah *Cluster* Optimum Pada Pengelompokan Provinsi Di Indonesia (Studi Kasus: Indikator Pendidikan Tingkat SMA/Sederajat)**". Penulisan skripsi ini bertujuan untuk memenuhi syarat memperoleh gelar Sarjana Statistika pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura.

Penulis menyadari dalam penulisan skripsi ini tidak luput dari bantuan, bimbingan, dukungan, serta motivasi dan doa dari berbagai pihak, sehingga penulis dapat menyelesaikan skripsi ini dengan lancar. Pada kesempatan ini, penulis ingin menyampaikan rasa terimakasih yang sebesar-besarnya terutama kepada:

1. Allah SWT. yang selalu melimpahkan segala berkah, rahmat, kemudahan, kelancaran, serta kekuatan bagi penulis dalam menyelesaikan skripsi ini.
2. Kedua orang tua penulis, Bapak Supriyanto dan Ibu Heru Nuryati, juga abang Iyan Satya Budi Darma yang selalu mendoakan, memberi dukungan, dan semangat yang luar biasa.
3. Ibu Shantika Martha, M.Si. selaku dosen pembimbing pertama yang telah mengarahkan dan membimbing penulis selama penulisan skripsi ini.
4. Ibu Neva Setyahadewi, S.Si, M.Sc, CRA, CRP, CRMP selaku dosen pembimbing kedua yang telah mengarahkan dan membimbing penulis selama penulisan skripsi ini.
5. Ibu Nurfitri Imro'ah, M.Si selaku dosen penguji pertama yang telah memberikan saran, masukan, serta koreksi demi penyempurnaan skripsi ini.
6. Bapak Hendra Perdana, M.Sc selaku dosen penguji kedua yang telah memberikan saran, masukan, serta koreksi demi penyempurnaan skripsi ini.
7. Bapak/Ibu Dosen Jurusan Matematika yang telah memberikan ilmu yang bermanfaat selama masa perkuliahan.

8. Sahabat-sahabat penulis, Hera, Luis, Della, Muthi, Ulya, Wahida, Siska, Nanda, dan Ghina yang selalu menemani, mewarnai, serta saling menyemangati selama masa perkuliahan.
9. Teman-teman Statistika angkatan 2021 “Expos” yang membersamai penulis selama menempuh pendidikan di Program studi Statistika FMIPA Untan.
10. Sahabat-sahabat penulis, Utin, Diana, Danuh, Farra, Egi, Adhis, Nisa, Ayu, Ipeh, Arta, Amanda, Azza, Nayla, Aboy, Farhan, Ravi, Budi, Angga, dan Yoga yang terus membersamai, mendukung dan menyemangati penulis sejak dulu.
11. Semua pihak yang telah membantu serta memberi dukungan kepada penulis, namun tidak dapat disebutkan satu persatu

Semoga segala hal yang telah diberikan kepada penulis menjadi amal ibadah serta diberkahi Allah SWT. Penulis mengharapkan saran serta kritik yang membangun untuk kebaikan dalam skripsi ini.

Pontianak, 23 Januari 2025

Putri Rahmadanti

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 3
1.3 Tujuan Penelitian 3
1.4 Batasan Masalah 3
1.5 Tinjauan Pustaka 4
1.6 Metodologi Penelitian 5
BAB II LANDASAN TEORI **8**
2.1 Pendidikan 8
2.2 Analisis Multivariat 8
2.3 Analisis *Cluster* 9
2.3.1 Metode Hierarki 10
2.3.2 Metode Non-Hierarki 11
2.4 Transformasi *Z-Score* 11
2.5 Ukuran Jarak atau Similaritas 12
BAB III PENENTUAN JUMLAH *CLUSTER* OPTIMUM **13**
3.1 Metode *Ward* 13
3.2 *Silhouette Coefficient* 13
BAB IV HASIL DAN PEMBAHASAN **16**
4.1 Deskripsi Data 16
4.2 Uji Multikolinearitas 17
4.3 Standarisasi Data 18

4.4 Perhitungan Jarak *Euclidean* 19
4.5 Pengelompokan Menggunakan Metode *Ward* 20
4.6 Penentuan Jumlah *Cluster* Optimum Menggunakan Metode *Silhouette Coefficient* 22
4.7 Interpretasi Profil *Cluster* 27
BAB V PENUTUP 31
5.1 Kesimpulan 31
5.2 Saran 31
DAFTAR PUSTAKA 33
LAMPIRAN 36

DAFTAR TABEL

Tabel 3.1 Nilai *Silhouette Coefficient* 15
Tabel 4.1 Variabel Penelitian 16
Tabel 4.2 Statistik Deskriptif 17
Tabel 4.3 Nilai *Variance Inflation Factor* 18
Tabel 4.4 Pengelompokan Provinsi dengan Dua *Cluster* 22
Tabel 4.5 Perbandingan Hasil Pengelompokan Menggunakan Metode *Silhouette Coefficient* 26
Tabel 4.6 Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Tingkat SMA/Sederajat 26
Tabel 4.7 Rata-Rata Variabel Indikator Pendidikan Tiap *Cluster* 27

DAFTAR GAMBAR

Gambar 1.1 *Flowchart* Penentuan Jumlah *Cluster* Optimum 7
Gambar 4.1 Statistik Deskriptif 17
Gambar 4.2 Dendogram Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Tingkat SMA/Sederajat 21
Gambar 4.3 Pemetaan Hasil *Cluster* Provinsi di Indonesia Berdasarkan Indikator Pendidikan Tingkat SMA/Sederajat 29

DAFTAR LAMPIRAN

Lampiran I Data Indikator Pendidikan Tingkat SMA/Sederajat Tiap Provinsi di Indonesia Tahun 2023 36
Lampiran II Data Hasil Standarisasi 37
Lampiran III Matriks Jarak *Euclidean* 38

DAFTAR SIMBOL

Z_i	: nilai *Z-Score* objek ke-*i*, dengan $i = 1, 2, 3,\ldots, n$
x_i	: nilai objek ke-*i*
$\bar{x}$	: nilai rata-rata
s	: nilai standar deviasi
n	: banyak objek
t	: banyak variabel
d_{ij}	: jarak *Euclidean* antar objek ke-*i* dengan objek ke-*j*, dengan $i = 1, 2, 3,\ldots, n$, dan $j = 1, 2, 3,\ldots, n$
x_{im}	: nilai objek ke-*i* pada variabel ke-*m*, dengan $m = 1, 2, 3,\ldots, t$
x_{jm}	: nilai objek ke-*j* pada variabel ke-*m*, dengan $m = 1, 2, 3,\ldots, t$
$d_{(PQ)R}$	: jarak antara *cluster* PQ dan R
$d_{(PR)}$	: jarak antara *cluster* P dan R
$d_{(QR)}$	: jarak antara *cluster* Q dan R
$d_{(PQ)}$	: jarak antara *cluster* P dan Q
$n_{(PQ)}$	: banyak objek pada *cluster* PQ
n_R	: banyak objek pada *cluster* R
a_i^k	: rata-rata jarak tiap objek ke-*i* dengan semua objek yang ada pada *cluster* yang sama
n_k	: banyak objek pada *cluster* ke-*k*
$d(x_i^k, x_l^k)$	: jarak objek ke-*i* dan objek ke-*l* pada *cluster k*
$d_i(k)$	: rata-rata jarak tiap objek ke-*i* dengan semua objek di *cluster* lainnya
k	: *cluster* ke-*k*, dengan $k = 1, 2, 3,\ldots, K$
n_z	: banyak objek pada *cluster* ke-*z*
$d(x_i^k, x_l^z)$	: jarak objek ke-*i* pada *cluster k* dan objek ke-*l* pada *cluster* lainnya

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara kepulauan yang terletak di Asia Tenggara. Indonesia terkenal dengan beragam suku bangsa, bahasa, agama, serta kebudayaan. Selain terkenal dengan segala keragamannya, Indonesia dikenal memiliki jumlah penduduk yang terbilang banyak. Jumlah penduduk Indonesia pada tahun 2023 tercatat 278,67 juta jiwa. Angka ini meningkat 2,9 juta jiwa dari tahun sebelumnya, yang tercatat berjumlah 275,77 juta jiwa. Jumlah penduduk yang akan meningkat setiap tahunnya menjadi tantangan besar bagi pemerintah untuk meningkatkan pembangunan nasional. Seiring berkembangnya teknologi, pengetahuan, serta banyak aspek lainnya, masih banyak didapatkan berbagai permasalahan di ranah sosial dan ekonomi yang menghalangi peningkatan pembangunan nasional. Oleh karena itu, perlu dikaji kembali mengenai penetapan kebijakan pemerintah dalam menyelesaikan masalah yang ada. Salah satu masalah besar yang sampai saat ini masih belum menemukan titik penyelesainnya adalah pendidikan (Rabiah, 2019).

Pendidikan merupakan aspek yang terkait dengan kualitas sumber daya manusia yang memiliki peran penting dalam upaya peningkatan pembangunan nasional. Seperti yang diketahui keadaan pendidikan di Indonesia akhir-akhir ini sangat memprihatinkan (Fitri, 2021). Hal ini disebabkan oleh sektor manajemen pendidikan yang rendah, kesenjangan fasilitas pendidikan di wilayah tertentu, kurikulum yang tidak sesuai, kemampuan para pengajar untuk menggali potensi murid, dan kurangnya kesejahteraan para pengajar. Pada dasarnya pemerataan pendidikan sangat krusial untuk meningkatkan pendidikan di Indonesia (Wahyudi, Mulyana, Dhiaz, Ghandari, Dinata, Fitoriq, & Hasyim, 2022). Pendidikan yang tidak merata ini menjadi salah satu alasan pendidikan di Indonesia terbilang sangat mengkhawatirkan dibandingkan dengan negara lain. Oleh sebab itu, perlu ditetapkan kebijakan dan upaya perubahan guna menunjang pemerataan kualitas pendidikan di tiap provinsi di Indonesia. Agar tujuan pemerataan pendidikan dapat tercapai, maka perlu strategi yang terencana dan tepat sasaran guna mencari solusi dari permasalahan tersebut. Maka dari itu, tentu perlu kerjasama yang baik antara

pemerintah dalam membuat kebijakan dengan pengajar serta murid yang menjadi sasaran kebijakan tersebut.

Masalah pendidikan terutama pada tingkat SMA/sederajat merupakan tantangan besar bagi pemerintah Indonesia dalam membangun sumber daya manusia yang nantinya berguna untuk meningkatkan pembangunan nasional. Hal ini karena lulusan SMA/sederajat akan berkontribusi langsung pada pembentukan tenaga kerja yang kemudian akan mencetak angkatan kerja dan akan mendukung peningkatan sumber daya manusia. Menurut BPS, Angka Partisipasi Murni (APM) tingkat SMA/sederajat, yaitu persentase jumlah siswa pada usia 16-18 tahun yang bersekolah pada jenjang pendidikan SMA/sederajat terhadap jumlah seluruh penduduk pada kelompok usia 16-18 tahun adalah sebesar 62,53%. Angka ini masih tergolong rendah dan membuat pemerintah harus kembali fokus dalam upaya peningkatan pendidikan agar dapat terus meningkat dari waktu ke waktu.

Indonesia yang terdiri dari banyak provinsi dengan beragam kondisi menjadikan perlu dilakukan pengelompokan terhadap pendidikan pada tiap daerah. Hal ini dilakukan guna menunjang kemudahan pemerintah Indonesia untuk memberi fokus lebih kepada daerah dengan pendidikan rendah agar dapat segera teratasi dengan maksimal dan tepat sasaran. Pendidikan dipengaruhi oleh banyak faktor (Fajri & Afriansyah, 2019), sehingga proses analisis kasus ini menggunakan analisis multivariat yang memungkinkan untuk memeriksa data dari beberapa variabel sekaligus. Pada penelitian kali ini digunakan analisis *cluster* yang merupakan salah satu metode pada analisis multivariat untuk menunjang proses pengelompokan.

Adapun fokus dari penelitian ini ialah menentukan jumlah *cluster* optimum menggunakan *Silhouette Coefficient* pada pengelompokan provinsi di Indonesia berdasarkan indikator pendidikan tingkat SMA/sederajat. Pengelompokan dilakukan dengan metode *Ward*, yaitu metode pengelompokan yang bersifat *agglomerative* (penggabungan) dengan tujuan memperoleh *cluster* dengan varian internal sekecil mungkin (Paramadina, Sudarmin, & Aidid, 2019). Dalam penentuan jumlah *cluster* optimum, digunakan *Silhouette Coefficient*, yang

merupakan salah satu metode yang digunakan untuk mengevaluasi keakuratan hasil analisis *cluster* yang diperoleh dengan mengukur berapa dekat relasi antara satu objek dengan objek lain pada satu *cluster* dan mengukur berapa jauh suatu *cluster* yang terpisah dengan *cluster* lainnya (Sholeh & Aeni, 2023).

1.2 Rumusan Masalah

Berlandaskan latar belakang yang telah dipaparkan, didapat rumusan masalah yaitu bagaimana penggunaan metode *Silhouette Coefficient* dalam menentukan jumlah *cluster* optimum pada pengelompokan provinsi di Indonesia berdasarkan indikator pendidikan tingkat SMA/sederajat?

1.3 Tujuan Penelitian

Berlandaskan permasalahan yang telah dipaparkan, tujuan dari penelitian ini yaitu menentukan jumlah *cluster* optimum menggunakan metode *Silhouette Coefficient* pada pengelompokan provinsi di Indonesia berdasarkan indikator pendidikan tingkat SMA/sederajat.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini yaitu:

1. Data yang digunakan adalah data sekunder yang didapatkan dari Badan Pusat Statistik (BPS), yaitu publikasi Statistik Pendidikan volume 12 tahun 2023.
2. Data yang digunakan merupakan data yang terkumpul per tahun 2023 dari setiap provinsi di Indonesia.
3. Pada penelitian ini hanya digunakan 11 variabel saja, yaitu Angka Partisipasi Kasar (APK) SMA/sederajat, Angka Partisipasi Murni (APM) SMA/sederajat, Rata-Rata Lama Sekolah (RLS) penduduk umur 15 tahun ke atas, angka melek aksara penduduk 15 tahun ke atas, tingkat penyelesaian pendidikan SMA/sederajat, jumlah Sekolah Menengah Atas (SMA), persentase ruang kelas baik pada jenjang SMA/sederajat, persentase guru yang memenuhi kualifikasi pada jenjang SMA/sederajat, rasio murid-guru SMA/sederajat, persentase

peserta didik umur 5-24 tahun yang mengakses internet selama 3 bulan terakhir, dan persentase peserta didik umur 5-24 tahun yang menggunakan komputer selama 3 bulan terakhir.

4. Penelitian ini menggunakan variabel persentase peserta didik umur 5-24 tahun yang mengakses internet selama 3 bulan terakhir dan persentase peserta didik umur 5-24 tahun yang menggunakan komputer selama 3 bulan terakhir karena dianggap memiliki relevansi yang lebih tinggi untuk jenjang SMA/sederajat, yaitu dilihat pada proses pembelajaran hingga tugas-tugas di tingkat SMA yang lebih kompleks dan memerlukan penggunaan komputer dan internet yang lebih intensif.
5. Perhitungan jarak untuk membentuk matriks menggunakan rumus jarak *Euclidean.*

1.5 Tinjauan Pustaka

Dewi, Talakua, Lesnussa, dan Matdoan (2021) melakukan pengelompokan kabupaten/kota di Provinsi Maluku yang dilakukan berdasarkan indikator pendidikan dengan menggunakan metode *Ward*. Penelitian ini menggunakan 6 variabel, yaitu banyak siswa SD, banyak siswa SMP, banyak siswa SMA/SMK, banyak guru SD, banyak guru SMP, dan banyak guru SMA/SMK. Pengelompokan dengan metode *Ward* ini dilakukan menggunakan jarak *Euclidean*. Hasil dari penelitian ini didapatkan bahwa *cluster* optimal memperoleh hasil 3 *cluster* dengan tingkatan, yaitu *cluster* pertama berisi kabupaten/kota dengan tingkat pendidikan tinggi, *cluster* kedua berisi kabupaten/kota dengan tingkat pendidikan rendah, dan *cluster* ketiga berisi kabupaten/kota dengan tingkat pendidikan sedang.

Muthiah dan Rosmaini (2023) melakukan pemetaan kemiskinan di Provinsi Sumatera Utara tahun 2019 menggunakan analisis *cluster hirarki* dengan metode *Ward* dan *Average Linkage.* Pemetaan kabupaten/kota ini didasarkan oleh 6 variabel yang mendukung indikator kemiskinan. Dari pemetaan 33 Kabupaten/Kota di Sumatera Utara berdasarkan indikator kemiskinan, metode *Ward* memberikan pemahaman yang lebih baik tentang karakteristik kemiskinan di tiap wilayah.

Iklima dan Pujiyanta (2023) melakukan perbandingan metode *K-Means Clustering* dan metode *Ward* dalam mengelompokkan pelangan mall yang kemudian dievaluasi menggunakan nilai *Silhouette*. Melalui perbandingan dua metode ini dihasilkan kesimpulan bahwa metode *Ward* lebih unggul dibandingkan metode *K-Means* dengan nilai *Silhouette* sebesar 0,81.

Sam, Syaripuddin, dan Wasono (2021) melakukan penelitian mengenai analisis *cluster* pada produk mie instan berdasarkan komposisi yang terkandung dengan menggunakan metode *Ward* menggunakan metode jarak *Euclidean*. Melalui penelitian ini diperoleh hasil bahwa nilai k optimum berada dititik 9 yang berarti jumlah *cluster* yang baik untuk digunakan sebanyak 9.

1.6 Metodologi Penelitian

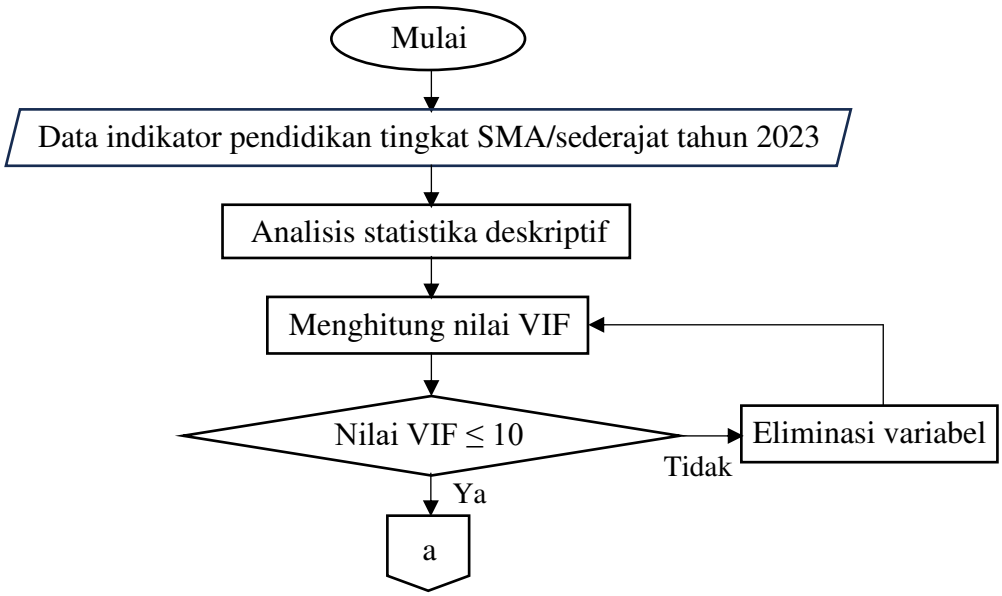
Penelitian ini menggunakan data sekunder yang didapatkan dari publikasi Badan Pusat Statistik (BPS). Jumlah observasi yang digunakan sesuai dengan jumlah provinsi di Indonesia yaitu sebanyak 34 provinsi dengan 11 variabel yang berkaitan dengan pendidikan. Adapun pengolahan data dilakukan dengan bantuan *software* R Studio.

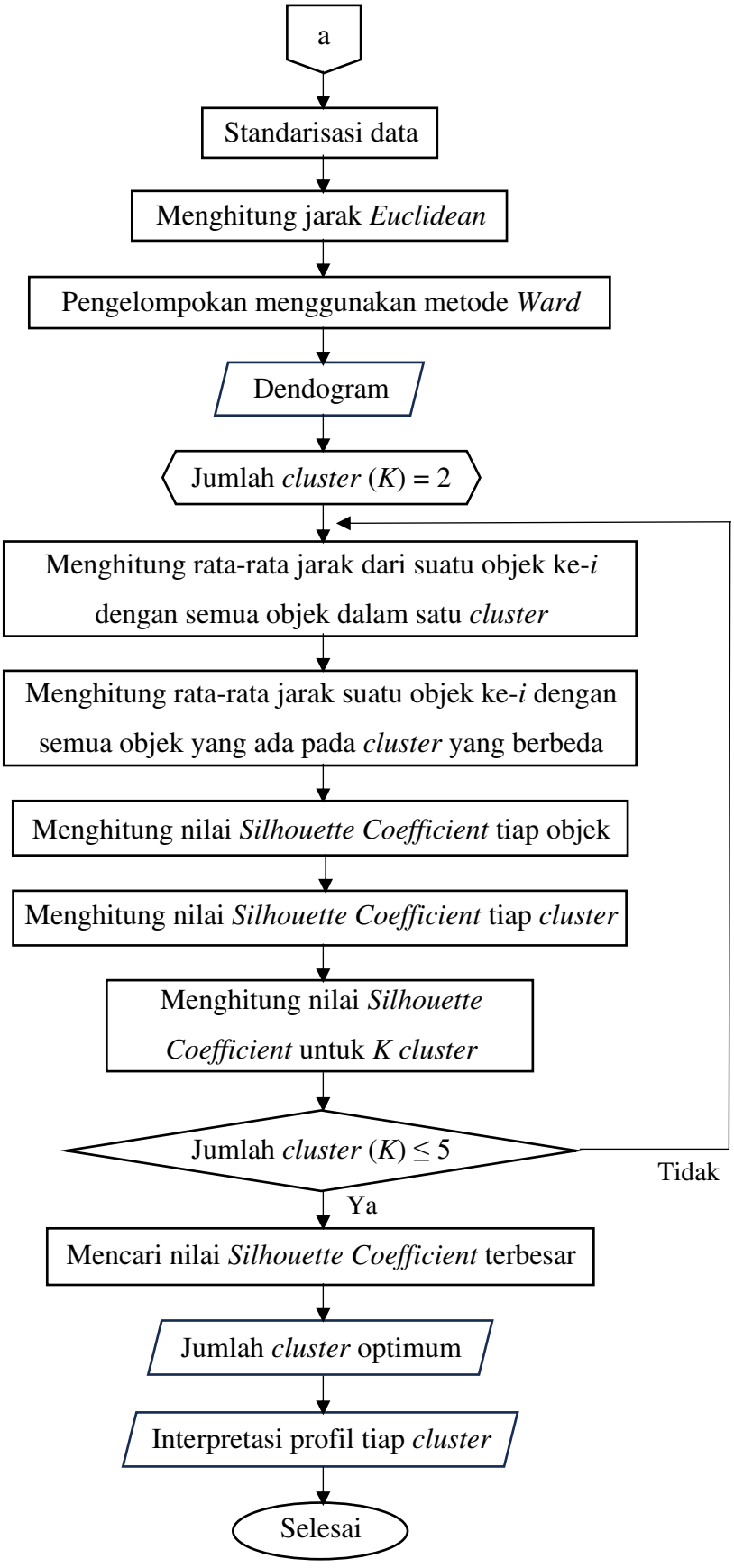
Tahapan pertama yang dilakukan untuk menentukan jumlah *cluster* optimum dalam pengelompokan provinsi di Indonesia berdasarkan indikator pendidikan tingkat SMA/sederajat yaitu melakukan analisis deskriptif terhadap data tiap indikator yang digunakan. Selanjutnya dilakukan pengujian multikolinearitas untuk melihat korelasi antar variabel dengan menggunakan nilai *Variation Inflation Factor* (VIF). Jika nilai VIF semua variabel kurang dari sama dengan 10, maka tidak terjadi multikolinearitas sehingga proses penelitian dapat diteruskan ke tahapan selanjutnya. Namun, jika terdapat variabel dengan nilai VIF kurang dari 10, maka perlu dilakukan eliminasi variabel dan kemudian dilanjutkan pada proses standarisasi data. Tahap standarisasi data dilakukan dengan menggunakan nilai *Z-Score*. Ini dilakukan karena data yang digunakan memiliki satuan yang bervariasi. Tahap berikutnya ialah menghitung ukuran kemiripan objek menggunakan jarak *Euclidean*. Kemudian dilakukan pengelompokan

menggunakan metode *Ward*. Hasil dari pengelompokan disajikan dalam bentuk dendogram.

Tahap selanjutnya dilakukan perhitungan rata-rata jarak dari suatu objek ke-*i* dengan semua objek yang ada pada satu *cluster* yang sama untuk masing-masing objek *i*. Dilanjutkan dengan menghitung rata-rata jarak suatu objek ke-*i* dengan semua objek yang ada pada *cluster* yang berbeda untuk masing-masing objek *i*, lalu pilih nilai yang paling kecil. Setelah itu dilakukan perhitungan nilai *Silhouette* untuk tiap objek. Kemudian hitung nilai *Silhouette* untuk tiap *cluster* dengan rata-rata nilai *Silhouette* dari tiap objek pada masing-masing *cluster*. Selanjutnya menghitung nilai *Silhouette Coefficient* untuk mengetahui kebaikan untuk *K cluster* yang terbentuk. Jika sebelumnya dilakukan perhitungan untuk *cluster* yang bukan berjumlah lima, maka dilanjutkan dengan perhitungan kembali untuk *cluster* yang berjumlah *K*+1 dengan tahap yang sama mulai dari menghitung rata-rata jarak tiap objek hingga menghitung nilai *Silhouette Coefficient*. Sedangkan jika perhitungan sebelumnya dilakukan untuk *cluster* yang berjumlah lima, maka dilanjutkan dengan mencari nilai *Silhouette Coefficient* terbesar dari beberapa nilai hasil perhitungan *Silhouette Coefficient* sebanyak *cluster* yang diperoleh sebelumnya. Setelah jumlah *cluster* optimum yang terbentuk telah diketahui, maka dilakukan tahap akhir yaitu menginterpretasikan profil masing-masing *cluster.*

Tahapan yang dilakukan dalam menentukan jumlah *cluster* optimum menggunakan metode *Silhouette Coefficient* ditampilkan pada Gambar 1.1 berikut.





Gambar 1.1 *Flowchart* Penentuan Jumlah *Cluster* Optimum