ANALISIS *CLUSTER HIERARKI* DENGAN METODE *AVERAGE LINKAGE* PADA INDIKATOR INDEKS PEMBANGUNAN MANUSIA DI KALIMANTAN BARAT

AMELIA VICKYLIA ROOS
H1011201061

SKRIPSI

PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2024

ANALISIS *CLUSTER HIERARKI* DENGAN METODE *AVERAGE LINKAGE* PADA INDIKATOR INDEKS PEMBANGUNAN MANUSIA DI KALIMANTAN BARAT

AMELIA VICKYLIA ROOS
H1011201061

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

PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2024

LEMBAR PENGESAHAN

Judul Tugas Akhir	:	Analisis *Cluster Hierarki* Dengan Metode *Average Linkage* Pada Indikator Indeks Pembangunan Manusia Di Kalimantan Barat
Nama Mahasiswa	:	Amelia Vickylia Roos
NIM	:	H1011201061
Jurusan/ Program Studi	:	Matematika/ Matematika
Tanggal Lulus	:	10 Desember 2024
SK Pembimbing	:	No.705/UN22.8/TD.06/2024/ Tanggal 1 Maret 2024
SK Penguji	:	No.4107/UN22.8/TD.06/2024/Tanggal 3 Desember 2024

Dosen Pembimbing

Pembimbing I

Nur'ainul Miftahul Huda, M.Si
NIP.199411142020122014

Pembimbing II

Drs. Helmi, M.Si
NIP.196410171998021001

Dosen Penguji

Ketua Penguji

Dr. Evi Noviani, M.Si.,CFrA
NIP.198402292006042001

Anggota Penguji

Asri Rahmawati, M.Mat
NIP.199509062024062001

Pimpinan Sidang
(merangkap anggota penguji)

Nur'ainul Miftahul Huda, M.Si
NIP.199411142020122014

Sekretaris Sidang
(merangkap anggota penguji)

Drs. Helmi, M.Si
NIP.196410171998021001

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	: Amelia Vickylia Roos
NIM	: H1011201061
Program Studi/ Jurusan	: 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 39Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pertanyaan ini untuk digunakan sebagaimana mestinya

Pontianak, 10 Desember 2024



Amelia Vickylia Roos
H1011202061

Analisis *Cluster Hierarki* Dengan Metode *Average Linkage* Pada Indikator Indeks Pembangunan Manusia Di Kalimantan Barat

ABSTRAK

Indeks Pembangunan Manusia (IPM) adalah indikator yang digunakan untuk mengukur dan membandingkan tingkat perkembangan manusia antar negara atau wilayah. Faktor-faktor yang mempengaruhi IPM sendiri yaitu, pengeluaran perkapita, rata-rata lama sekolah, harapan lama sekolah, umur harapan hidup saat lahir, umur harapan hidup, dan presentasi penduduk miskin. Untuk melakukan pengelompokkan metode yang paling sering digunakan adalah metode analisis *cluster*. Metode *cluster hierarki* adalah metode yang sering digunakan untuk mengelompokkan wilayah atau negara. *Average lingkage* adalah teknik statistik yang mengelompokkan data berdasarkan rata-rata kesamaan karakteristik, Proses ini dimulai dengan mencari faktor yang mempengaruhi IPM kemudian dilakukan standarisasi karena memliki satuan yang berbeda. Di lakukan uji multikolinearitas untuk, menentukan jarak *Euclidean* untuk menentukan seberapa dekat jarak satu titik ke titik lainnya. Proses *clustering* pada analisis hierarki dapat ditentukan jumlah clusternya. Pada penelitian ini yang digunakan adalah 3 klaster. Hasil yang didapat adalah kota Pontianak dan Kayong Utara mempunyai karakteristik yang berbeda dengan kabupaten/kota lainnya disebabkan oleh faktor-faktor IPM yang lebih unggul. Untuk 12 Kabupaten/kota lainnya memiliki karakteristik yang hampir mirip sehingga menjadi satu *cluster*.
Kata kunci: Indeks Pembangunan Manusia, *Cluster Hierarki*.

Hierarchical Cluster Analysis With Average Linkage Method On Human Development Indicators In West Kalimantan

ABSTRACT

The Human Development Index (HDI) is an indicator used to measure and compare the level of human development between countries or regions. Factors that affect the HDI itself are per capita expenditure, average length of schooling, expected length of schooling, life expectancy at birth, life expectancy, and percentage of poor population. To do the grouping, the most frequently used method is the cluster analysis method. The hierarchical cluster method is a method often used to group regions or countries. Average linkage is a statistical technique that groups data based on the average similarity of characteristics. This process begins by looking for factors that affect the HDI and then standardization is carried out because they have different units. A multicollinearity test is carried out to determine the Euclidean distance to determine how close one point is to another. The clustering process in hierarchical analysis can determine the number of clusters. In this study, 3 clusters were used. The results obtained were that the cities of Pontianak and Kayong Utara have different characteristics from other districts/cities due to superior HDI factors. For 12 other districts/cities, they have almost similar characteristics so that they become one cluster.
Keywords: Human Development Index, hierarchical cluster.

PRAKATA

Segala puji dan syukur kami panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat dan berkat-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "ANALISIS *CLUSTER HIERARKI* DENGAN METODE *AVERAGE LINKAGE* PADA INDIKATOR INDEKS PEMBANGUNAN MANUSIA DI KALIMANTAN BARAT" ini tepat pada waktunya. Dalam menyelesai studi dan penulisan skripsi ini, penulis banyak memperoleh bantuan baik pengajaran, bimbingan dan arahan dari berbagai pihak baik secara langsung maupun tidak langsung. untuk itu penulis menyampaikan terima kasih yang tak terhingga kepada :

1. Ibu Nur'ainul Miftahul Huda, M.Si selaku dosen pembimbing pertama saya yang telah memberikan arahan serta bimbingan dalam penulisan skripsi ini.
2. Drs. Helmi, M.Si selaku dosen Pembimbing kedua saya yang telah memberikan arahan serta bimbingan dalam penulisan skripsi ini.
3. Dr. Evi Noviani, M.si CFrA dan Ibu Asri Rahmawati, M.Mat selaku dosen penguji yang memberikan nasehat dalam menyelesaikan skripsi ini.
4. Orang tua penulis untuk beliau berdualah skripsi ini saya persembahkan. terima kasih atas segala kasih sayang serta dukungan sehingga penulis bisa menyelesaikan skripsi ini.
5. Fabio Cloudio selaku pacar, teman, sahabat, serta calon penulis, terima kasih telah banyak meluangkan baik waktu, tenaga, dan materi kepada saya.
6. Putri Ananda, Alma putri awalia, Dea, April, Eka, Fanny, Ayu, Elma, serta teman-teman penulis yang tidak bisa di sebutkan satu persatu.

Akhir kata, penulis mengharapkan skripsi ini dapat memberikan manfaat bagi penulis khususnya dan bagi pembaca pada umumnya.

Penulis

Amelia Vickylia Roos

DAFTAR ISI

LEMBAR PENGESAHAN ii
PERNYATAAN INTEGRITAS AKADEMIK iii
ABSTRAK iv
ABSTRACT v
PRAKATA vi
DAFTAR ISI vii
DAFTAR TABEL ix
DAFTAR GAMBAR x
DAFTAR SIMBOL xii
DAFTAR LAMPIRAN 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 3
1.6 Metodologi Penelitian 5
BAB II LANDASAN TEORI 7
2.1 Analisis *Cluster* 7
2.1 Standarisasi Data 11
2.3 Jarak Eucliden 12
2.3 Uji Asumsi Analisis *Cluster* 13
2.4 Indeks Pembangunan Manusia 14
BAB III HASIL DAN PEMBAHASAN 23
3.1 Deskripsi Data 23
3.2 Hasil Statistik Deskriptif Variabel Penelitian 25
3.3 Standarisasi Data 27
3.4 Asumsi Analisis *Cluster* 27

3.5 Analisis *Cluster* 30
BAB IV KESIMPULAN 33
DAFTAR PUSTAKA 34
LAMPIRAN 37

DAFTAR TABEL

Tabel 3.1 Data 23
Tabel 3.2 Statistik Deskriptif 25
Tabel 3.3 Uji Multikolinieritas 28

DAFTAR GAMBAR

Gambar 2.1 Contoh Dendogram 8
Gambar 2.2 Perkembangan IPM Kalimantan Barat 16
Gambar 2.3 IPM Kalimantan Barat 18
Gambar 2.4 Pengeluaran Perkapita Kalimantan Barat 19
Gambar 2.5 Rata-rata Lama Sekolah Kalimantan Barat 20
Gambar 2.6 Harapan Lama Sekolah Kalimantan Barat 21
Gambar 2.7 Umur Harapan Hidup saat lahir Kalimantan Barat 22
Gambar 2.8 Umur Harapan Hidup Kalimantan Barat 23
Gambar 2.9 Persentasi Penduduk Miskin Kalimantan Barat 23
Gambar 3.1 Hasil *Dendogram* 30
Gambar 3.2 Hasil *Clustering* 31

DAFTAR SIMBOL

x_i : data ke-i

$\bar{x}$: rata-rata data

Z : *Z-score*

s : simpangan baku

$d_{i,k}$: jarak antara objek i dengan objek-k

y_{ik} : nilai objek i pada variabel ke-k

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

p : banyaknya variabel *cluster*

n : banyaknya pengamatan

X_{i1} : nilai pengamatan ke-i pada variabel X_1

X_{i2} : nilai pengamatan ke-i pada variabel X_2

$d_{i,k}$: jarak antara objek i dalam kelompok (UV) dan objek k dalam kelompok W

$N_{(UV)}$: banyaknya objek-objek dalam kelompok (UV)

$N_{(w)}$: banyaknya objek-objek dalam kelompok W

DAFTAR LAMPIRAN

Lampiran 1 Data Standarisasi 37
Lampiran 2 Jarak *Euclidean* 38

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pembangunan manusia merupakan suatu proses yang bertujuan untuk meningkatkan pilihan yang tersedia bagi masyarakat. Di antara banyak pilihan tersebut, yang paling penting adalah panjang umur dan sehat, memiliki pengetahuan dan akses terhadap sumber daya yang diperlukan untuk menjalani kehidupan yang layak (UNDP, 2024).

Indeks Pembangunan Manusia (IPM) adalah ukuran yang digunakan untuk menilai perkembangan sosial dan ekonomi suatu negara atau wilayah. IPM dibangun menggunakan pendekatan dasar tiga dimensi. Aspek-aspek tersebut antara lain panjang umur dan sehat, pengetahuan, dan kehidupan yang layak. Ketiga dimensi ini mempunyai arti yang sangat luas karena berkaitan dengan banyak faktor. Untuk mengukur aspek kesehatan digunakan rata-rata angka harapan hidup saat lahir. Selain itu, untuk mengukur aspek pengetahuan digunakan kombinasi angka tidak buta huruf dan rata-rata lama sekolah. Sementara itu, untuk mengukur aspek-aspek kehidupan yang layak masyarakat mengunakan indikator daya beli masyarakat terhadap kebutuhan dasar yang dihitung dari pengeluaran perkapita sebagai pendekatan pendapatan yang mewakili pencapaian perkembangan kehidupan yang layak (BPS, 2020).

Badan Pusat Statistik (BPS) mencatat bahwa indeks pembangunan manusia (IPM) Kalimantan Barat pada 2023 masuk ke tingkat tinggi dengan nilai 70,74% meningkat (1,09) dibanding tahun sebelumnya (69,71). Selama dua tahun belakangan status Pembangunan manusia di Kalbar berada di Tingkat sedang sedangkan pada tahun 2023 sudah berada di Tingkat tinggi. Penyusunan IPM Kalimantan Barat 2023 bertujuan untuk beberapa hal utama: pertama, memperkenalkan metode baru dalam perhitungan IPM; kedua, memantau kemajuan pembangunan manusia di provinsi Kalimantan Barat; ketiga, menyediakan gambaran yang lebih sederhana dan menyeluruh tentang dampak pembangunan yang telah dilakukan serta implikasinya terhadap peningkatan

kualitas hidup masyarakat; dan keempat, memberikan laporan tahunan mengenai perkembangan IPM di setiap kabupaten/kota beserta kontribusi mereka dalam mempercepat pencapaian IPM Kalimantan Barat. Informasi ini diharapkan dapat membantu berbagai pihak dalam merancang program dan kebijakan di provinsi Kalimantan Barat, terutama yang terkait dengan pembangunan manusia (BPS , 2020).

Dalam ilmu statistika, ketika variabel yang digunakan lebih dari dua, analisis statistik menjadi lebih kompleks. Salah satu metode untuk mengatasi hal ini adalah analisis multivariat. Analisis multivariat merupakan teknik statistik yang memungkinkan penelitian terhadap lebih dari dua variabel. Salah satu teknik dalam analisis multivariat adalah analisis *cluster* (Hair J & Anderson 2014). Analisis *cluster* adalah cabang dari ilmu statistika yang digunakan untuk mengelompokkan objek-objek tertentu berdasarkan kesamaan karakteristik menjadi beberapa kelompok. Proses pengelompokan ini dapat menggunakan dua metode, yaitu *cluster* hierarki dan *cluster* non-hierarki. *Cluster hierarki* adalah metode pengelompokan yang mengorganisir objek ke dalam struktur pohon berdasarkan kemiripan atau jarak antara mereka. Proses ini dimulai dengan mengelompokkan dua objek yang paling mirip, lalu secara bertahap menambahkan objek lain ke dalam kelompok atau membentuk kelompok baru, hingga semua objek tergabung dalam satu kelompok besar. Metode ini dapat dibagi menjadi dua pendekatan utama: penggabungan (a*gglomerative*) dan pembagian (*divisive*) (Han & Tong, 2022).

Penelitian sebelumnya yang berjudul "Pengelompokan Provinsi Berdasarkan Prioritas Potensi Sektor Maritim Indikator Blue Economy menggunakan Analisis *Cluster Average Linkage*" oleh Koesnadi dan ana tahun 2023 membahas tentang analisis *cluster* dan menemukan tiga cluster yang optimal. Oleh karena itu, penelitian ini menggunakan metode analisis *cluster hierarki* untuk indeks pembangunan manusia tahun 2023, yang akan menghasilkan kelompok-kelompok berbeda dengan karakteristik masing-masing. Analisis *cluster* ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi indeks pembangunan manusia tahun 2023.

Berdasarkan uraian tersebut, penulis menyusun penelitian dengan judul "Analisis *Cluster Hierarki* dengan Metode *Average Linkage* Pada Indikator Indeks Pembangunan Manusia di Kalimantan Barat." Penelitian ini diharapkan dapat menambah pengetahuan serta informasi mengenai faktor-faktor yang mempengaruhi indeks pembangunan manusia pada tahun 2023.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah dalam penelitian ini, yaitu bagaimana menghasilkan pengelompokkan kabupaten/kota di Kalimantan Barat tahun 2023 dengan metode analisis *cluster hirarki average lingkage*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dalam penelitian ini untuk mengelompokkan kabupaten/kota di Kalimantan Barat tahun 2023 menggunakan analisis *cluster hirarki average lingkage* berdasarkan faktor-faktor indeks pembangunan manusia.

1.4 Batasan Masalah

Adapun Batasan masalah yang ada dalam penelitian adalah

1. Metode yang dipakai dalam pembentukan *cluster* adalah metode *cluster hirarki* dengan metode *Average Lingkage*.
2. Dalam melakukan analisis peneliti di lakukan standarisasi data karena data yang digunakan memiliki satuan yang berbeda.
3. Data yang digunakan adalah data sekunder yang didapat pada setiap kabupaten/kota yang ada di Kalimantan Barat pada Badan Pusat Statistik tahun 2023.
4. Dalam melakukan analisis ini peneliti hanya mengunnakan enam faktor yang mempengaruhi IPM yaitu, pengeluaran perkapita, rata-rata lama sekolah, harapan lama sekolah, umur harapan hidup saat lahir, umur harapan hidup dan persentase penduduk miskin.

1.5 Tinjauan Pustaka

Penelitian terdahulu mengenai *Average Lingkage* yang dilakukan oleh Faran dan Aldisa pada tahun 2023 dengan judul “Analisis Data Mining dalam Komparasi *Average Lingkage* AHC dan K-Means Clustering untuk Dataset

Facebook Live Seller". Hasil dari penelitian ini didapat bahwa kedua metode tersebut menghasilkan tiga *cluster* dengan jumlah data yang sama. *Cluster* 1 mencakup 336 data, yang mewakili 98,5% dari total data yang dikelompokkan. *Cluster* 2 terdiri dari 2 item, yang hanya menyumbang 0,6% dari total data. Sedangkan *Cluste*r 3 memiliki 3 item, dengan persentase 0,9% dari keseluruhan data yang dikelompokkan

Dalam penelitian yang menggunakan analisis *cluster* Ningsih dan Silvia 2016 mengelompokkan kabupaten/kota di Provinsi Kalimantan Timur berdasarkan data produksi palawija. Mereka menggunakan metode *complete linkage* dan *average linkage* dalam pendekatannya. hasil penelitian tersebut menghasilkan empat kelompok kabupaten/kota: kelompok pertama merupakan kabupaten/kota dengan produksi palawija yang sangat rendah, kelompok kedua memiliki produksi palawija yang cukup tinggi, kelompok ketiga adalah penghasil palawija terbanyak, dan kelompok keempat kembali menunjukkan kabupaten/kota dengan produksi palawija yang sangat rendah.

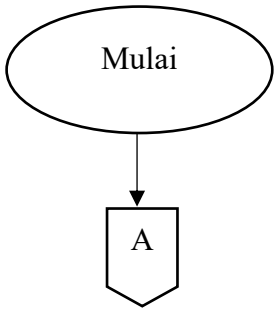
Menurut penelitian Alwi dan Hasrul pada tahun 2018 diperoleh bahwa hasil pengelompokkan Kabupaten/Kota di provinsi Sulawesi Selatan berdasarkan indikator kesejahteraan rakyat menggunakan metode Average Linkage di dapat tiga *cluster* yaitu *cluster* pertama terdiri dari 21 kabupaten/kota, dengan indikator utama variabel Kepemilikan Rumah Sendiri (X11). Selain variabel PDRB (X1) dan Kepemilikan Rumah Sendiri (X11), tujuh variabel lainnya memiliki rata-rata nilai centroid terendah dibandingkan cluster lainnya. *Cluster* kedua di dapat dua kabupaten/kota, dengan indikator utama variabel Rata-Rata Lama Sekolah. Pada Cluster 2, variabel PDRB (X1) memiliki rata-rata yang paling rendah, sedangkan delapan variabel lainnya memiliki rata-rata nilai centroid yang cukup tinggi (sedang). dan *cluster* terakhir didapat satu kabupaten/kota, yaitu Kota Makassar, dengan variabel utama Kepadatan Penduduk (X2). Pada cluster ini, delapan variabel lainnya memiliki rata-rata nilai centroid tertinggi dibandingkan cluster lainnya, sedangkan variabel Kepemilikan Rumah Sendiri (X11) memiliki rata-rata nilai centroid paling rendah.

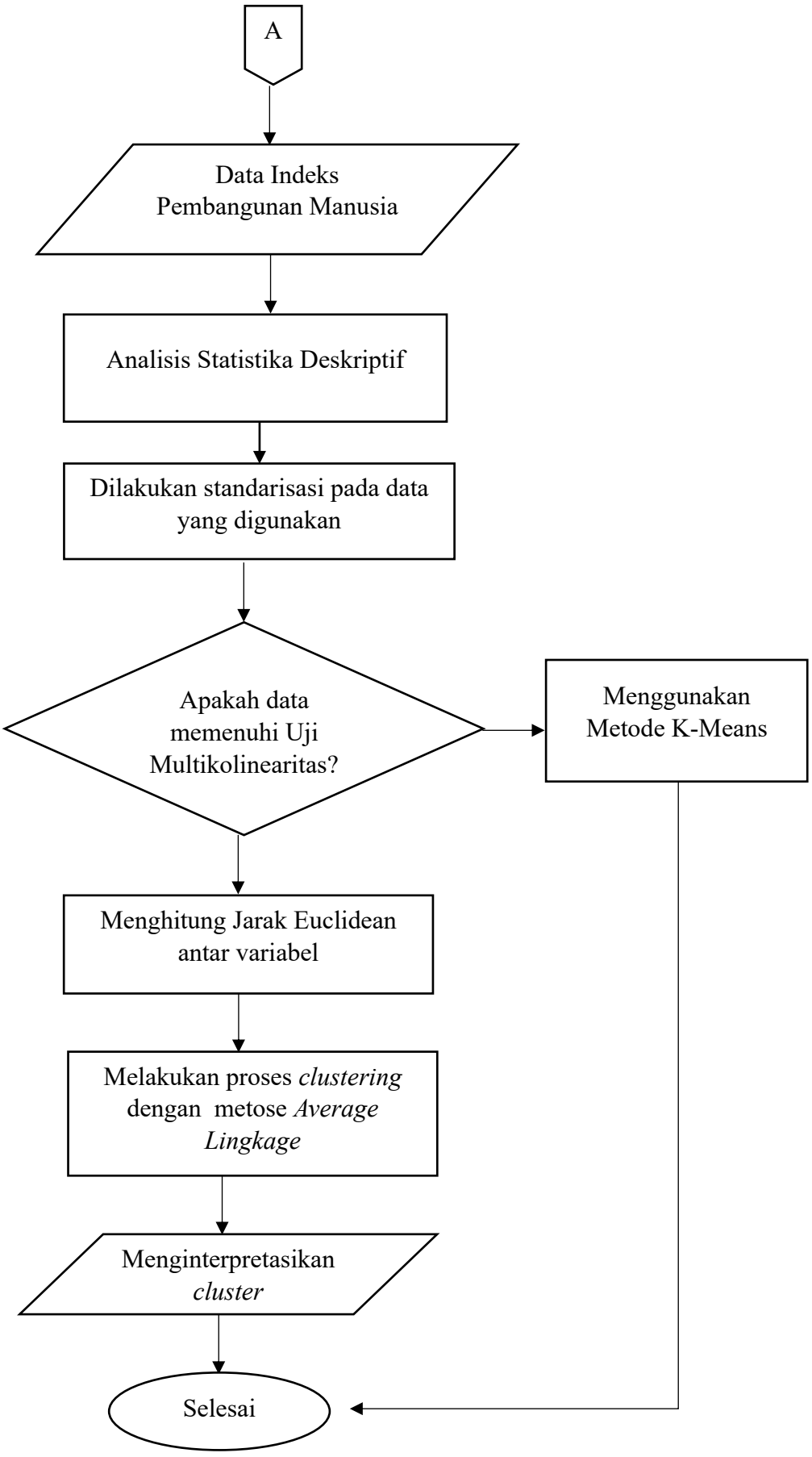
Menurut Abdurrahman pada tahun 2019 dengan judul "*Clustering* Data Kredit Bank Menggunakan *Agglomerative Hierarchical Clustering Average Linkage*" mendapat kesimpulan bahwa Algoritma clustering *agglomerative hierarchical* dengan metode *average linkage* adalah teknik *unsupervised learning* yang menggabungkan dua cluster untuk membentuk cluster baru, berdasarkan kesamaan yang diukur dengan jarak Euclidean, menggunakan rata-rata jarak antar individu dalam cluster lain. Dari pengujian data sebanyak 822 (100%), ditemukan 3 cluster: cluster-1 dengan 806 anggota (98%), cluster-2 dengan 5 anggota (1%), dan cluster-3 dengan 9 anggota (1%). Cluster-4 dan cluster-5 masing-masing hanya memiliki 1 anggota (0%).

1.6 Metodologi Penelitian

Penelitian ini diawali dengan mencari referensi yang relevan dengan topik yang akan diteliti. Data yang digunakan merupakan data sekunder yang di dapat dari badan pusat statistik Kalimantan Barat pada tahun 2023. Teknik analisis data digunakan pada penelitian ini adalah analisis multivariat dengan metode *average* dilakukan dengan banntuan *software* R studio dan *microsoft Excel*.

Pertama mengumpulkan data-data yang mempengaruhi indeks pembangunan manusia di Kalimantan Barat tahun 2023. Kemudian dilakukan statistika deskriptif pada data yang akan digunakan. Kemudian di lakukan standarisasi data karena satuan data yang digunakan berbeda. Lalu, di lakukan uji multikorelinearitas untuk melihat apakah variabel yang digunakan layak atau tidak. menentukan jarak *Euclidean* yang untuk menentukan seberapa dekat atau jauh satu titik data dari titik data lainnya. Menarik kesimpulan yang didapat mengunakan dengan mengunakan metode *average lingkage*.




A
Data Indeks Pembangunan Manusia
Analisis Statistika Deskriptif
Dilakukan standarisasi pada data yang digunakan
Apakah data memenuhi Uji Multikolinearitas?
Menggunakan Metode K-Means
Menghitung Jarak Euclidean antar variabel
Melakukan proses *clustering* dengan metose *Average Lingkage*
Menginterpretasikan *cluster*
Selesai