PENGELOMPOKAN DAERAH DI KALIMANTAN BARAT BERDASARKAN INDIKATOR TINGKAT PENGANGGURAN DENGAN *HIERARCHICAL CLUSTERING MULTISCALE BOOTSTRAP*

NANDA HAFIFAH
NIM H1091211064

SKRIPSI

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

PENGELOMPOKAN DAERAH DI KALIMANTAN BARAT BERDASARKAN INDIKATOR TINGKAT PENGANGGURAN DENGAN *HIERARCHICAL CLUSTERING MULTISCALE BOOTSRAP*

NANDA HAFIFAH
NIM H1091211064

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	:	**Pengelompokan Daerah di Kalimantan Barat Berdasarkan Indikator Tingkat Pengangguran dengan *Hierarchical Clustering Multiscale Bootstrap***
Nama Mahasiswa	:	Nanda Hafifah
NIM	:	H1091211064
Jurusan/Program Studi	:	Matematika/Statistika
Tanggal Lulus	:	13 Juni 2025
SK Pembimbing	:	2638/UN22.8/TD.06/2024/Tanggal 2 September 2024
Sk Penguji	:	1696/UN22.8/TD.06/2025/Tanggal 5 Juni 2025

Dosen Pembimbing

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

Dosen Penguji

Ketua Penguji	Anggota Penguji
Nurfitri Imro'ah, M.Si	Yudhi, M.Si
NIP. 198907182019032021	NIP. 198504072019031004

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Hendra Perdana, M.Sc	Dr. Evy Sulistianingsih, M.Sc
NIP. 198810102019031020	NIP. 198502172008122006

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	: Nanda Hafifah
NIM	: H1091211064
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, 13 Juni 2025



Nanda Hafifah

NIM. H1091211064

Pengelompokan Daerah di Kalimantan Barat Berdasarkan Indikator Tingkat Pengangguran dengan *Hierarchical Clustering Multiscale Bootstrap*

Abstrak

Pengangguran merupakan masalah umum di negara berkembang, termasuk Indonesia, yang dipengaruhi oleh beberapa faktor seperti pertumbuhan penduduk, ekonomi, ketersediaan lapangan kerja, tingkat pendidikan, serta besaran upah. Tingginya populasi di Indonesia menjadi tantangan dalam menekan angka pengangguran. Untuk mengukur dampaknya dalam suatu wilayah, digunakan indikator tingkat pengangguran, yaitu persentase jumlah pengangguran dibandingkan dengan total angkatan kerja. Berdasarkan data BPS tahun 2023, Kalimantan Barat mencatat tingkat pengangguran sebesar 5,05% yang masih tergolong tinggi. Penelitian ini bertujuan untuk mengelompokkan Kabupaten/Kota di Kalimantan Barat berdasarkan indikator tingkat pengangguran dengan *Hierarchical Clustering Multiscale Bootstrap* serta mengevaluasi validitas setiap *cluster*. Analisis *cluster* adalah teknik analisis multivariat yang digunakan untuk mengelompokkan objek berdasarkan kesamaan karakteristik. Salah satu metode hierarki yang digunakan adalah metode Ward, yang meminimalkan varians internal dalam *cluster*. Validasi dilakukan dengan metode *resampling multiscale bootstrap*, yang menghasilkan dua nilai, yaitu *bootstrap probability* (BP) dan *approximately unbiased* (AU). Nilai BP menunjukkan seberapa sering suatu cluster terbentuk saat data diulang, sedangkan AU dianggap lebih akurat dalam menilai keandalan *cluster*. Sebuah *cluster* dianggap stabil jika nilai AU $\geq$ 0,95. Namun, dalam penelitian ini digunakan batas minimal AU $\geq$ 0,75. Hasilnya, diperoleh dua *cluster* yang stabil yaitu *cluster* 1 dengan tingkat pengangguran rendah terdiri dari Mempawah, Kayong Utara, Sanggau, Kota Singkawang, Landak, Sambas, Sintang, Bengkayang, Kapuas Hulu, Sekadau, dan Melawi. Cluster 3 dengan tingkat pengangguran tinggi yaitu Kabupaten Kubu Raya dan Kota Pontianak. Serta *cluster* 2 dengan tingkat pengangguran sedang yaitu Ketapang, namun memiliki nilai AU sebesar 0,72.

Kata kunci: Metode *Ward*, *Multiscale Bootstrap*, Tingkat Pengangguran

Classification of Regions in West Kalimantan Based on Unemployment Rate Indicators Using Hierarchical Clustering Multiscale Bootstrap

Abstract

Unemployment is a common issue in developing countries, including Indonesia, influenced by several factors such as population growth, economic conditions, job availability, education level, and wage rates. Indonesia's large population presents a challenge in reducing unemployment. To measure its impact in a region, the unemployment rate is used, defined as the percentage of unemployed individuals in the total labor force. According to BPS data in 2023, West Kalimantan recorded an unemployment rate of 5.05%, which is considered relatively high. This study aims to cluster regencies/cities in West Kalimantan based on unemployment indicators using Hierarchical Clustering with Multiscale Bootstrap and to evaluate the validity of each cluster. Cluster analysis is a multivariate technique used to group objects based on similarities in characteristics. One of the hierarchical methods used is Ward's method, which minimizes within-cluster variance. Validation was conducted using the multiscale bootstrap resampling method, which produces two values: bootstrap probability (BP) and approximately unbiased (AU). BP indicates how frequently a cluster appears when the data is resampled, while AU is considered more accurate in assessing cluster reliability. A cluster is regarded as stable if AU ≥ 0.95; however, this study uses a minimum threshold of AU ≥ 0.75. The results show three clusters: Cluster 1 (low unemployment) includes Mempawah, Kayong Utara, Sanggau, Singkawang City, Landak, Sambas, Sintang, Bengkayang, Kapuas Hulu, Sekadau, and Melawi; Cluster 3 (high unemployment) includes Kubu Raya and Pontianak City, and Cluster 2 (moderate unemployment) consists of Ketapang, but it has an AU value of 0.72, below the minimum threshold.

Keywords**:** *Ward Method*, *Multiscale Bootstrap*, *Unemployment*

PRAKATA

Puji syukur kehadirat Allah SWT karena atas limpahan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penulisan skripsi ini. Skripsi ini berjudul **"Pengelompokan Daerah di Kalimantan Barat Berdasarkan Indikator Tingkat Pengangguran dengan *Hierarchical Clustering Multiscale Bootstrap*"**. Skripsi ini dibuat dengan tujuan untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Statistika pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura Pontianak.

Dengan selesainya penulisan skripsi ini, penulis menyadari bahwa selama proses penulisan skripsi ini tidak lepas dari bimbingan, arahan, dan dukungan dari berbagai pihak. Maka dari itu penulis ingin mengucapkan terima kasih kepada:

1. Kedua orang tua, kakak, abang beserta keluarga besar tercinta yang selalu senantiasa memberi semangat, nasihat, motivasi, dan mendoakan tanpa pernah putus. Serta dukungan secara moral dan materi kepada penulis selama masa perkuliahan maupun penulisan skripsi ini.
2. Bapak Hendra Perdana, M.Sc selaku Dosen Pembimbing Pertama yang telah meluangkan waktu, memberikan bimbingan, dukungan, serta saran dalam penulisan tugas akhir ini.
3. Ibu Dr. Evy Sulistianingsih, M.Sc selaku Dosen Pembimbing Kedua yang telah meluangkan waktu, memberikan bimbingan, dan saran dalam penulisan tugas akhir ini.
4. Ibu Nurfitri Imro'ah, M.Si selaku Dosen Penguji Pertama yang telah memberikan masukan, saran, serta kritikan kepada penulis.
5. Bapak Yudhi, M.Si selaku Dosen Penguji Kedua sekaligus Pembimbing Akademik yang telah meluangkan waktu, memberikan bimbingan, masukan, serta saran selama perkuliahan kepada penulis.

6. Bapak dan Ibu Dosen Pengajar di Program Studi Statistika yang telah memberikan pengetahuan, pengalaman, serta ilmu dalam proses pembelajaran semasa perkuliahan.
7. Della, Fransiska, Ghina, Hera, Luis, Muthiara, Putri, Ulya, Wahida, dan Arya yang selalu memberikan motivasi, dukungan dan doa, yang selalu menjadi pendengar dan menghibur penulis, serta melalui segala hal selama masa perkuliahan bersama-sama baik senang maupun susah.
8. Teman-teman statistika 2021 "Expos" yang telah membersamai dan berjuang bersama selama masa perkuliahan.
9. Seluruh pihak yang tidak dapat disebutkan satu pesatu yang telah memberikan dukungan serta doa kepada penulis dalam penyelesaian tugas akhir ini.

Pontianak, 13 Juni 2025



Nanda Hafifah

DAFTAR ISI

Abstrak ..... **i**

Abstract ..... **ii**

PRAKATA ..... **iii**

DAFTAR ISI ..... **v**

DAFTAR GAMBAR ..... **vii**

DAFTAR TABEL ..... **vii**

DAFTAR LAMPIRAN ..... **ix**

DAFTAR SIMBOL ..... **x**

BAB I PENDAHULUAN **1**

1.1 Latar belakang 1

1.2 Rumusan Masalah 2

1.3 Tujuan Penelitian 3

1.4 Batasan Masalah 3

1.5 Tinjauan Pustaka 3

1.6 Metodologi Penelitian 4

BAB II LANDASAN TEORI **7**

2.1 Indikator Tingkat Pengangguran 7

2.2 Analisis Multivariat 8

2.3 Analisis *Cluster* 9

2.4 Metode Analisis *Cluster* 12

2.5 Validasi *Cluster* 14

BAB III *MULTISCALE BOOTSTRAP* **15**

3.1 Metode *Ward* 15

3.2 *Bootstrap* 16

3.3 *Multiscale Bootstrap* 16
BAB IV HASIL DAN PEMBAHASAN 19
4.1 Deskripsi Data 19
4.2 Analisis Statistik Deskriptif 19
4.3 Uji Multikolinearitas 20
4.4 Standarisasi Data 21
4.5 Analisis *Cluster* 21
4.6 Metode *Multiscale Bootstrap* 26
BAB V KESIMPULAN 31
DAFTAR PUSTAKA 32
LAMPIRAN 35

DAFTAR GAMBAR

Gambar 1. 1 Diagram Alir *Hierarchical Clustering Multiscale Bootstrap*..... 6
Gambar 4. 1 Dendogram Analisis *Cluster*..... 23
Gambar 4. 2 Peta Hasil Pengelompokan Indikator Tingkat Pengangguran di Kalimantan Barat..... 26
Gambar 4. 3 Plot Diagnostik dari *Multiscale Bootstrap* untuk Langkah Penggabungan 1..... 27
Gambar 4. 4 Dendogram *Multiscale Bootstrap*..... 29

DAFTAR TABEL

Tabel 4.1 Variabel Penelitian 19
Tabel 4.2 Analisis Statistik Deskriptif..... 20
Tabel 4.3 Nilai VIF 20
Tabel 4.4 Anggota Hasil *Cluster* 24
Tabel 4.5 Rata-rata Setiap Variabel pada *Cluster* 24
Tabel 4.6 Hasil Perhitungan AU, BP, v dan c dengan Metode *Multiscale Bootstrap* 28

DAFTAR LAMPIRAN

Lampiran I Data Indikator Tingkat Pengangguran di Kabupaten/Kota Provinsi Kalimantan Barat Tahun 2023.....35
Lampiran II Hasil Standarisasi Data 35
Lampiran III Tabel Jarak *Euclidean* 36
Lampiran IV Nilai AU 36

DAFTAR SIMBOL

R_k^2	: nilai koefisien determinasi variabel dependen dan independen
$Z_{i,k}$	: nilai standarisasi objek ke-*i* pada variabel ke-*k*
μ_k	: rata-rata variabel ke-*k*
σ_k	: nilai standar deviasi ke-*k*.
d_{ij}	: jarak *euclidean* dari objek ke-*i* dan objek ke-*j*
p	: jumlah variabel *cluster*
x_{ik}	: nilai objek ke-*i* pada variabel ke-*k*
x_{jk}	: nilai objek ke-*j* pada variabel ke-*k*.
$d_{(UW)}$	: jarak antara *cluster U* dan *W*
$d_{(VW)}$	: jarak antara *cluster V* dan *W*
$d_{(UV)W}$	: jarak antara *cluster UV* dan *W*
$n_{(UV)}$	: jumlah objek pada *cluster UV*
n_W	: jumlah objek pada *cluster W*
N	: jumlah data pengamatan
N'	: jumlah data pengamatan *bootstrap*
m	: jumlah iterasi
$z(N')$	: nilai *z* sebagai fungsi dari ukuran sampel *bootstrap* N'
AU	: nilai *Appoximately Unbiased*
Φ	: fungsi distribusi normal
v dan c	: estimasi parameter untuk AU

BAB I
PENDAHULUAN

1.1 Latar belakang

Pengangguran adalah salah satu masalah yang sering muncul di sektor ekonomi negara-negara berkembang, termasuk di Indonesia. Pengangguran dapat disebabkan oleh beberapa faktor, seperti laju pertumbuhan yang tinggi, laju pertumbuhan ekonomi, ketersediaan lapangan kerja yang tidak merata, tingkat pendidikan yang rendah, serta besaran upah (Suhadi & Setyowati, 2022). Populasi penduduk Indonesia yang tinggi menjadi tantangan bagi pemerintah dalam menangani masalah pengangguran. Berdasarkan data Badan Pusat Statistik tahun 2023 tingkat pengangguran di Indonesia tercatat 5,32%. Persentase ini mengalami penurunan sebesar 0,54% jika dibandingkan dengan tingkat pengangguran pada Agustus 2022 yaitu 5,86% (BPS, 2024).

Tingkat pengangguran di Provinsi Kalimantan Barat tercatat sebesar 5,05% dimana mengalami penurunan sebesar 0,06% dibandingkan tahun 2022. Meskipun demikian, angka tersebut masih dalam kategori yang tinggi (BPS, 2024). Untuk mengukur dampaknya dalam suatu wilayah, digunakan indikator tingkat pengangguran, yaitu persentase jumlah pengangguran dibandingkan terhadap total angkatan kerja. Tinggi rendahnya tingkat pengangguran dapat memengaruhi tingkat kemiskinan di suatu wilayah, karena pengangguran merupakan salah satu faktor utama terhadap tingkat kemiskinan (David, Engka, & Sumual, 2019). Oleh karena itu, diperlukan penelitian terkait dengan tingkat pengangguran beserta indikator yang berpengaruh, yaitu dengan menggunakan metode pengelompokan, seperti analisis *cluster*.

Analisis *cluster* merupakan analisis statistik yang bertujuan untuk mengelompokkan objek-objek berdasarkan kesamaan/kemiripan karakteristik diantara objek tersebut (Anuraga, 2015). Analisis *cluster* terbagi menjadi dua metode yaitu metode hierarki (*hierarchical clustering method)* dan non-hierarki (*non-hierarchical clustering method)*. Metode hierarki memiliki dua jenis pendekatan, yaitu *agglomerative* dan *divisive*. *Agglomerative* mempunyai

dua jenis pendekatan seperti *single linkage, average linkage, complete linkage,* dan *ward*. Sedangkan pada metode non-hierarki mempunyai pendekatan *K-means* (Supranto, 2004).

Salah satu kelemahan yang dimiliki analisis *cluster* yaitu tidak dapat menunjukkan hasil validasi dari hasil pengelompokan sehingga diperlukan metode lain untuk mengetahui hasil validasi dari analisis *cluster* yang telah dilakukan. Metode *multiscale bootstrap* merupakan metode yang dapat menunjukkan validitas dari masing-masing *cluster* yang terbentuk dari analisis sebelumnya. Metode ini bekerja dengan menghitung nilai *Approximately Unbiased* (AU) melalui pendekatan *bootstrap resampling* untuk setiap *cluster* yang terbentuk (Safarina, Hayati, dan Nasution, 2019). *Hierarchical clustering multiscale bootstrap* (HCMB) mengintegrasikan teknik *bootstrap* dengan algoritma *hierarchical clustering* dalam menghasilkan estimasi yang lebih akurat untuk struktur *clustering* dalam data dan stabilitas *cluster* (Shimodaira, 2004). Analisis HCMB mengevaluasi stabilitas hasil *hierarchical clustering* terhadap variasi dalam data dengan melakukan pengulangan *bootstrap* pada subset data dengan skala yang berbeda. Analisis ini dapat membantu dalam mengukur sejauh mana hasil *cluster* dapat dipercaya dan dipertahankan jika data diambil ulang atau jika variasi dalam data berubah (Kalamiyah, 2023)

Berdasarkan penjelasan di atas, pada penelitian ini dilakukan pengelompokan tingkat pengangguran yang terdapat di Kalimantan Barat pada tahun 2023 dengan analisis *hierarchical clustering* dengan metode *multiscale bootstrap* sebagai validasinya.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah pada penelitian ini adalah bagaimana hasil pengelompokan Kabupaten/Kota dengan *hierarchical clustering multiscale bootstrap* berdasarkan tingkat pengangguran serta mengetahui hasil stabilitas setiap *cluster?*

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat maka tujuan penelitian yang ingin dicapai dalam penelitian ini adalah mengkaji hasil pengelompokan Kabupaten/Kota dengan *hierarchical clustering multiscale bootstrap* berdasarkan tingkat pengangguran serta mengetahui hasil stabilitas setiap *cluster*.

1.4 Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini adalah:

1. Menggunakan data indikator tingkat pengangguran Kabupaten/Kota di Kalimantan Barat tahun 2023.
2. Menggunakan pengukuran jarak *euclidean.*
3. Menggunakan metode hierarki *clustering*.

1.5 Tinjauan Pustaka

Musdalifah, Sudarmin, dan Bobby (2022) melakukan penelitian mengenai Analisis *Hierarchical Clustering Multiscale Bootstrap* dengan kasus indikator kemiskinan di Provinsi Sulawesi Selatan tahun 2020. Analisis tersebut menggunakan metode *single linkage, average linkage,* dan *complete linkage.* Hasil penelitian dengan menggunakan ketiga metode tersebut adalah empat *cluster* dengan karakteristik yang berbeda. Diperoleh hasil bahwa pada *single linkage* terdapat empat *cluster* yang valid, *complete linkage* hanya satu cluster yang valid dan *average linkage* terdapat tiga *cluster* yang valid. Untuk itu, *single linkage* merupakan metode terbaik dalam mengelompokkan kabupaten/kota di Sulawesi Selatan berdasarkan indikator kemiskinan.

Penelitian oleh Yulia Safarina dkk. (2019) menggunakan *Hierarchical Clustering Multiscale Bootstrap* dengan metode *complete linkage* dalam mengelompokkan Kabupaten/Kota di Kalimantan Timur berdasarkan Indikator IPM tahun 2017. Hasil penelitian yang diperoleh yaitu terdapat empat *cluster* yang terbentuk. Kota Balikpapan, Kota Samarinda, dan Kota Bontang termasuk dalam *cluster* pertama, Kabupaten Kutai Kartanegara termasuk dalam *cluster* kedua, Kabupaten Penajam Paser Utara, Kabupaten Paser,

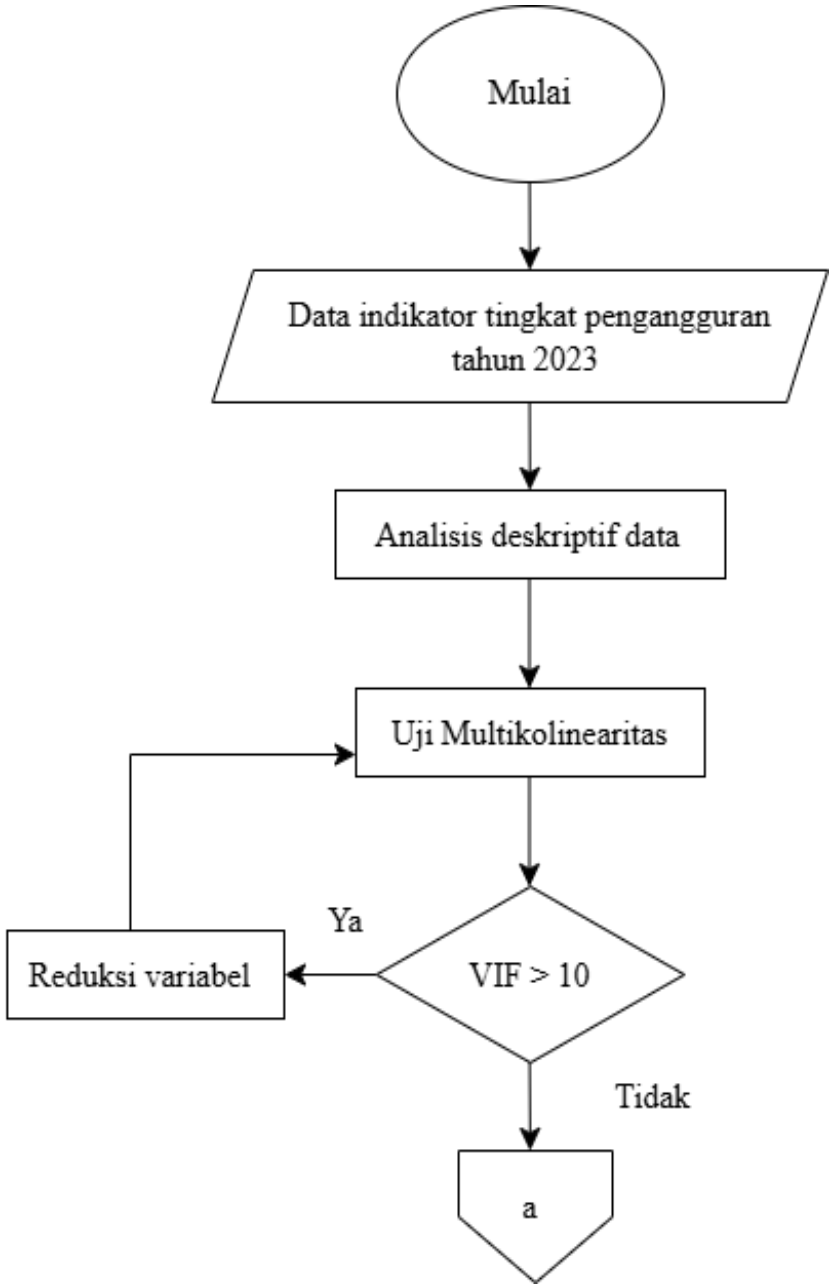
Kabupaten Kutai Barat, Kabupaten Kutai Timur, dan Kabupaten Berau termasuk dalam *cluster* ketiga, Kabupaten Mahakam Ulu termasuk dalam *cluster* keempat. Dari keempat *cluster* yang terbentuk, terdapat tiga *cluster* yang valid yaitu *cluster* kedua, *cluster* ketiga, dan *cluster* keempat.

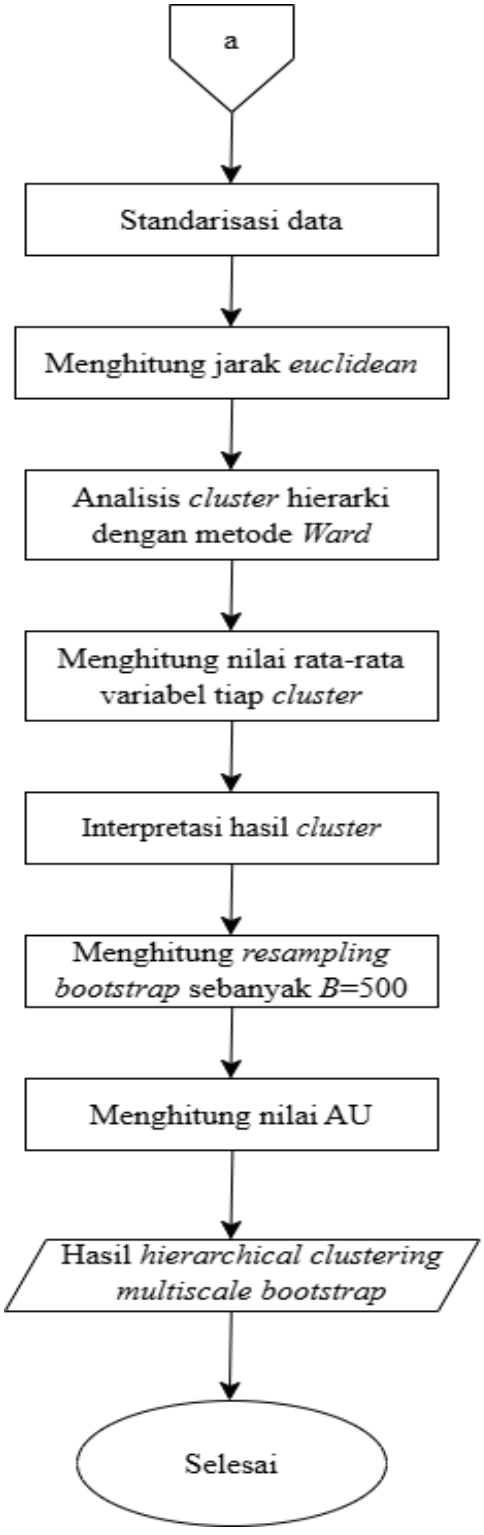
Penelitian yang dilakukan Anuraga (2015) mengenai *Hierarchical Clustering Multiscale Bootstrap* dalam mengelompokkan kemiskinan di Jawa Timur dengan menggunakan metode hierarki *complete linkage*. Pendekatan *multiscale bootstrap* dalam *Hierarchical Cluster Analysis* memiliki kestabilan pada saat iterasi B = 500. Hal itu dapat diketahui dari signifikansi nilai AU (*p-value*) yang mendekati 0.95 dengan nilai standar *error bootstrap* yang kecil. Terdapat lima kelompok (*cluster*) yang terbentuk dengan kemiripan satu dengan yang lain dan signifikan dengan nilai AU (*Approximately Unbiased*) ≥ 0,95. Berdasarkan signifikansi *Approximately unbiassed* (AU) *p-value* ≥ 0,95, kelompok kelima yang terdiri dari Kabupaten Jember, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan dapat dikatakan sebagai daerah yang masih cukup tinggi persentase kemiskinannya dibandingkan dengan kelompok 1, 2, 3, dan 4.

1.6 Metodologi Penelitian

Tahap pertama yang dilakukan dalam mengelompokkan daerah di Kalimantan Barat berdasarkan indikator tingkat pengangguran menggunakan metode *Ward* dengan validasi *Multiscale Bootstrap* adalah melakukan analisis deskriptif data pada tiap indikator yang digunakan. Selanjutnya, melakukan standarisasi data menggunakan *z-score* dikarenakan data yang digunakan memiliki satuan yang bervariasi. Setelah data di standarisasi, kemudian menghitung ukuran jarak antar objek menggunakan jarak *euclidean*. Berikutnya, dilakukan pengelompokan menggunakan analisis *cluster* hierarki dengan metode *Ward*. Setelah didapat hasil *cluster,* kemudian menghitung nilai rata-rata tiap variabel *cluster* yang selanjutnya di interpretasikan. Langkah berikutnya adalah menghitung nilai validitas menggunakan *Multiscale Bootstrap* untuk mengetahui validitas tiap *cluster* yang terbentuk.

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





Gambar 1. 1 Diagram Alir *Hierarchical Clustering Multiscale Bootstrap*