K-HARMONIC MEANS CLUSTERING DENGAN VALIDASI *DUNN INDEX* UNTUK PENGELOMPOKAN TINGKAT PENGANGGURAN DI KALIMANTAN BARAT

TIARA MARGARETA
H1091221014

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2026

K-HARMONIC MEANS CLUSTERING DENGAN VALIDASI *DUNN INDEX* UNTUK PENGELOMPOKAN TINGKAT PENGANGGURAN DI KALIMANTAN BARAT

TIARA MARGARETA
H1091221014

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
2026

LEMBAR PENGESAHAN

Judul Tugas Akhir	: *K-Harmonic Means Clustering* dengan Validasi *Dunn Index* untuk Pengelompokan Tingkat Pengangguran di Kalimantan Barat
Nama Mahasiswa	: Tiara Margareta
NIM	: H1091221014
Jurusan/Program Studi	: Matematika/Statistika
Tanggal Lulus	: 22 Mei 2026
SK Pembimbing	: No. 3571/UN22.8/TD.06/2025/Tanggal 9 Oktober 2025
SK Penguji	: No. 2350/DST/UN22.8/DT.00.09/2026/Tanggal 20 Mei 2026

Dosen Pembimbing

Pembimbing I	Pembimbing II
Nurfitri Imro'ah, M.Si.	Hendra Perdana, M.Sc.
NIP. 198907182019032021	NIP. 198810102019031020

Dosen Penguji

Ketua Penguji	Anggota Penguji
Wirda Andani, M.Si.	Shantika Martha, M.Si.
NIP. 199411152022032016	NIP. 198403082008122003

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Nurfitri Imro'ah, M.Si.	Hendra Perdana, M.Sc.
NIP. 198907182019032021	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	: Tiara Margareta
NIM	: H1091221014
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, 22 Mei 2026



Tiara Margareta
NIM H1091221014

K-Harmonic Means Clustering dengan Validasi *Dunn Index* untuk Pengelompokan Tingkat Pengangguran di Kalimantan Barat

Abstrak

Tingkat pengangguran merupakan salah satu permasalahan ketenagakerjaan yang dipengaruhi oleh berbagai faktor sosial dan ekonomi yang berbeda antar wilayah. Berdasarkan kondisi tersebut, diperlukan pengelompokan kabupaten/kota di Provinsi Kalimantan Barat berdasarkan faktor penyebab tingkat pengangguran menggunakan analisis *cluster*. Analisis *cluster* merupakan teknik pengelompokan yang bertujuan untuk mengelompokkan objek berdasarkan kemiripan karakteristik yang dimiliki. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Kalimantan Barat berdasarkan faktor penyebab tingkat pengangguran menggunakan metode *K-Harmonic Means* serta ditentukan jumlah *cluster* optimal menggunakan *Dunn Index*. Variabel yang digunakan dalam penelitian ini meliputi Tingkat Partisipasi Angkatan Kerja (TPAK), laju pertumbuhan Produk Domestik Regional Bruto (PDRB), persentase penduduk miskin, Rata-rata Lama Sekolah (RLS), angka ketergantungan, jumlah penduduk, serta realisasi investasi Penanaman Modal Asing (PMA) dan Penanaman Modal Dalam Negeri (PMDN). Berdasarkan hasil analisis, jumlah *cluster* optimal yang terbentuk adalah empat *cluster* dengan parameter $p = 2{,}5;\ 3;\ 3{,}5$; dan 4, dengan nilai *Dunn Index* sebesar 0,635. Hasil pengelompokan menunjukkan bahwa *Cluster* 1 memiliki TPAK yang relatif tinggi, namun disertai pertumbuhan PDRB terendah dan persentase penduduk miskin yang masih tinggi, sehingga penyerapan tenaga kerja belum optimal. *Cluster* 2 memiliki pertumbuhan PDRB dan realisasi PMDN yang tinggi, yang mencerminkan kondisi ekonomi dan pasar kerja yang lebih kondusif. *Cluster* 3 ditandai oleh jumlah penduduk dan realisasi PMA yang sangat tinggi, namun juga memiliki tingkat kemiskinan tertinggi, sehingga menunjukkan adanya ketimpangan kesejahteraan. Sementara itu, *Cluster* 4 memiliki RLS yang tinggi tetapi TPAK relatif rendah, yang mengindikasikan pemanfaatan tenaga kerja terdidik belum optimal.

Kata Kunci: *Cluster* Non-Hierarki, Jarak *Manhattan*, Validasi Internal, Pengangguran

K-Harmonic Means Clustering with Dunn Index Validation for Unemployment Rate Clustering in West Kalimantan

Abstract

The unemployment rate is one of the labor issues influenced by various social and economic factors that differ across regions. Based on this condition, it is necessary to cluster regencies/cities in West Kalimantan Province according to the factors influencing unemployment using cluster analysis. Cluster analysis is a grouping technique aimed at grouping objects based on their similarity in characteristics. This study aims to cluster regencies/cities in West Kalimantan Province based on the factors influencing unemployment using the K-Harmonic Means method, with the optimal number of clusters determined using the Dunn Index. The variables used in this study include the Labor Force Participation Rate (LFPR), Gross Regional Domestic Product (GRDP) growth, poverty rate, Mean Years of Schooling (MYS), dependency ratio, population, as well as Foreign Direct Investment (FDI) and Domestic Direct Investment (DDI). Based on the analysis results, the optimal number of clusters formed is four clusters with parameter $p = 2.5$; 3; 3.5; and 4, with a Dunn Index value of 0.635. The clustering results indicate that Cluster 1 is characterized by relatively high LFPR, but low GRDP growth and a high poverty rate, indicating suboptimal labor absorption. Cluster 2 with high GRDP growth and DDI, reflecting more favorable economic and labor market conditions. Cluster 3 is characterized by a very high population and FDI, but also the highest poverty rate, indicating inequality in welfare distribution. Meanwhile, Cluster 4 has a high MYS but relatively low LFPR, indicating that educated labor is not optimally utilized.

Keywords: *Non-Hierarchical cluster, Manhattan Distance, Internal Validation, Unemployment*

PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas kasih dan penyertaan-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "***K-Harmonic Means Clustering* dengan Validasi *Dunn Index* untuk Pengelompokan Tingkat Pengangguran di Kalimantan Barat**". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura.

Penulis menyadari bahwa dalam proses penyusunan skripsi ini tidak terlepas dari bimbingan, arahan, motivasi, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Kedua orang tua tercinta yaitu Bapak dan Mama, yang senantiasa memberikan doa, kasih sayang, dukungan moral maupun materi, serta motivasi yang tidak pernah putus kepada penulis selama menempuh pendidikan hingga penyusunan skripsi ini.
2. Kakak Erin serta adik Nova, Vaola, dan Samuel, yang senantiasa memberikan perhatian, dukungan, serta dorongan semangat kepada penulis selama proses penyusunan skripsi ini.
3. Ibu Nurfitri Imro'ah, M.Si. selaku Dosen Pembimbing Pertama, yang telah meluangkan waktu untuk memberikan bimbingan, arahan, saran, serta masukan, disertai dukungan dan motivasi kepada penulis sehingga tugas akhir ini dapat diselesaikan dengan baik.
4. Bapak Hendra Perdana, M.Sc. selaku Dosen Pembimbing Kedua, yang telah meluangkan waktu untuk memberikan bimbingan, arahan, saran, serta masukan, disertai dukungan dan motivasi kepada penulis sehingga tugas akhir ini dapat diselesaikan dengan baik.
5. Ibu Wirda Andani, M.Si selaku Dosen Penguji Pertama yang telah memberikan saran, masukan, serta koreksi demi penyempurnaan penulisan skripsi ini.

6. Ibu Shantika Martha, M.Si selaku Dosen Penguji Kedua yang telah memberikan saran, masukan, serta koreksi demi penyempurnaan penulisan skripsi ini.
7. Bapak dan Ibu Dosen Program Studi Statistika yang senantiasa memberikan ilmu pengetahuan, bimbingan, dan pengalaman berharga kepada penulis selama masa studi.
8. Teman-teman selama masa perkuliahan, Hazwani, Putri, dan Sari yang selalu menemani penulis serta menjadi pendengar dalam setiap perjalanan perkuliahan, baik saat senang maupun susah.
9. Teman-teman Majestic Statistika Angkatan 2022 yang telah berjuang bersama, saling mendukung, dan menemani penulis sejak awal perkuliahan hingga selesai.
10. Seluruh pihak yang telah memberikan bantuan, dukungan, dan kontribusi dalam berbagai bentuk selama proses penyusunan skripsi ini, yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa dalam penulisan tugas akhir ini masih ada kekurangan dan jauh dari sempurna. Oleh karena itu, saran dan masukan yang membangun dari pembaca sangat penulis harapkan. Semoga tugas akhir ini dapat menambah wawasan dan pengetahuan, serta bermanfaat bagi pembaca maupun pihak-pihak yang membutuhkan informasi terkait masalah yang sama.

Pontianak, 22 Mei 2026



Tiara Margareta
NIM H1091221014

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 4
1.4 Batasan Masalah 4
1.5 Tinjauan Pustaka 4
1.6 Metodologi Penelitian 6
BAB II LANDASAN TEORI 10
2.1 Analisis *Cluster* 10
2.2 *K-Means* 12
2.3 Standarisasi Data 13
2.4 Jarak Antar Objek 13
2.5 Tingkat Pengangguran 14
BAB III *K-HARMONIC MEANS* 16
3.1 *K-Harmonic Means* 16
3.2 Validasi Data Hasil *Clustering* 18
BAB IV HASIL DAN PEMBAHASAN 20
4.1 Deskripsi Data 20
4.2 Statistik Deskriptif 20
4.3 Deteksi *Outlier* 21
4.4 Uji Multikolinearitas 22
4.5 Standarisasi Data 23

4.6 Pengelompokan Menggunakan Metode *K-Harmonic Means* 23
4.6.1 Analisis Hasil Pengelompokan Menggunakan Metode *K-Harmonic Means* dengan 2 *Cluster* 23
4.6.2 Analisis Hasil Pengelompokan Menggunakan Metode *K-Harmonic Means* dengan 3 *Cluster* dan 4 *Cluster* 31
4.7 Validasi Data Hasil *Clustering* 32
BAB V KESIMPULAN 38
5.1 Kesimpulan 38
5.2 Saran 39
DAFTAR PUSTAKA 40
LAMPIRAN 44

DAFTAR TABEL

Tabel 4.1 Variabel Penelitian 20
Tabel 4.2 Statistik Deskriptif..... 21
Tabel 4.3 Hasil Uji VIF 22
Tabel 4.4 Nilai Fungsi Tujuan Pada $K = 2$ dan $p = 2; 2{,}5; 3; 3{,}5;$ dan 4 24
Tabel 4.5 Nilai Keanggotaan Data dengan Jumlah *Cluster* $(K) = 2$..... 26
Tabel 4.6 Nilai Bobot Data untuk $K = 2$ dengan $p = 2; 2{,}5; 3; 3{,}5;$ dan 4..... 27
Tabel 4.7 Titik Pusat *Cluster* Baru yang Diperoleh dengan $p = 2;\ 2{,}5;\ 3;\ 3{,}5;$ dan 4..... 29
Tabel 4.8 Hasil Penetapan Keanggotaan Data 2 *Cluster* dengan $p = 2; 2{,}5; 3; 3{,}5;$ dan 4..... 30
Tabel 4.9 Jumlah Kabupaten/Kota pada Setiap *Cluster* Berdasarkan Nilai p..... 31
Tabel 4.10 Nilai *Dunn Index* Hasil *Clustering*..... 34
Tabel 4.11 Hasil Pengelompokan 4 *Cluster* dengan $p = 2{,}5;\ 3;\ 3{,}5;$ dan 4 Menggunakan Metode *K-Harmonic Means*..... 35
Tabel 4.12 Nilai Rata-Rata Setiap Variabel Untuk Masing-Masing *Cluster*..... 35

DAFTAR GAMBAR

Gambar 1.1 *Flowchart* Penelitian..... 9
Gambar 4.1 *Boxplot* Variabel PDRB..... 22

DAFTAR LAMPIRAN

Lampiran 1 Data Penelitian 44
Lampiran 2 *Boxplot* Variabel Penelitian 45
Lampiran 3 Data Standarisasi 46
Lampiran 4 Hasil Perhitungan Jarak Antar *Cluster* 47
Lampiran 5 Hasil Perhitungan Jarak Dalam Setiap *Cluster* 48

DAFTAR SIMBOL

N	: Banyak pengamatan
K	: Banyak *cluster*
x_i	: Data pengamatan ke-i, $i = 1,2, \dots, N$
X	: Variabel pengamatan
m	: Banyak variabel
$\hat{x}_{i,j}$	: Standarisasi data ke-i pada variabel ke-j, $j = 1,2, \dots, m$
μ_j	: Nilai rata-rata dari variabel ke-j
σ_j	: Simpangan baku dari variabel ke-j
c_l	: Titik pusat *cluster* ke-l, $l = 1,2, \dots, K$
p	: Parameter pembobot dalam *K-Harmonic Means*
$\|x_i - c_l\|$	: Jarak antara data ke-i ke pusat *cluster* ke-l
$KHM(X, C)$	: Nilai fungsi tujuan metode *K-Harmonic Means* berdasarkan variabel pengamatan terhadap titik pusat *cluster*
$m(c_l \mid x_i)$	: Nilai keanggotaan titik pusat *cluster* ke-l terhadap data pengamatan ke-i
$w(x_i)$	: Nilai bobot untuk data pengamatan ke-i
$d(c_p, c_q)$	: Jarak minimum dari seluruh pasangan objek yang berasal dari kedua *cluster* p dan q
$diam\ (c_r)$	: Jarak maksimum antara seluruh pasangan objek yang berada dalam *cluster* ke-r
R^2	: Koefisien determinasi
VIF	: *Variance Inflation Vector*
DI	: *Dunn Index*

BAB I
PENDAHULUAN

1.1 Latar Belakang

Laju pertumbuhan penduduk yang semakin cepat menyebabkan Indonesia menghadapi berbagai permasalahan yang hingga kini belum dapat diselesaikan sepenuhnya. Salah satu permasalahan utama yang masih menjadi perhatian adalah pengangguran. Pengangguran merupakan permasalahan penting dalam pembangunan ekonomi suatu daerah, karena tingginya tingkat pengangguran mencerminkan kurang optimalnya pembangunan akibat ketidakseimbangan antara angkatan kerja dengan ketersediaan lapangan pekerjaan (Ranti, Astrid, Yanti, dan Barella, 2024). Data Badan Pusat Statistik mencatat bahwa pada Agustus 2024 jumlah penduduk pengangguran di Indonesia mencapai 7,47 juta orang, dengan proporsi 4,91 persen terhadap total angkatan kerja.

Tingkat pengangguran juga menjadi salah satu permasalahan utama yang dihadapi Provinsi Kalimantan Barat. Kondisi ketenagakerjaan di Provinsi Kalimantan Barat menunjukkan bahwa tingkat pengangguran pada Agustus 2024 mencapai 4,86 persen, mengalami penurunan sebesar 0,19 persen poin dibandingkan Agustus 2023. Namun, angka tersebut masih berada pada kisaran tingkat pengangguran nasional dan dibarengi dengan kenaikan jumlah pekerja setengah penganggur sebesar 1,47 persen, yang mengindikasikan bahwa kualitas lapangan kerja yang tersedia belum mampu menyerap tenaga kerja secara optimal serta adanya gejala *jobless growth* di mana pertumbuhan ekonomi belum linear dengan kesejahteraan buruh (Badan Pusat Statistik, 2024). Tingkat pengangguran berkaitan erat dengan kondisi ekonomi masyarakat, khususnya dalam hal ketersediaan lapangan pekerjaan dan tingkat pendapatan. Keterbatasan kesempatan kerja memengaruhi kondisi ketenagakerjaan serta kualitas pekerjaan yang tersedia (Sihombing, 2022).

Permasalahan pengangguran di Kalimantan Barat tidak hanya disebabkan oleh terbatasnya lapangan pekerjaan, tetapi juga dipengaruhi struktur ekonomi daerah yang masih didominasi sektor primer, keterbatasan kesempatan kerja formal, serta

ketidaksesuaian antara kualitas tenaga kerja dengan kebutuhan pasar kerja (Putri, Wijaya, Sofiah, Aini, dan Ulum, 2024). Selain itu, realisasi investasi di Kalimantan Barat belum sepenuhnya mampu mendorong pemerataan penyerapan tenaga kerja karena cenderung terkonsentrasi pada wilayah dan sektor tertentu seperti pertambangan dan perkebunan yang padat modal namun rendah dalam penyerapan tenaga kerja lokal. Kondisi ini turut diperkuat oleh adanya kesenjangan pembangunan antar wilayah, yang tercermin dari perbedaan kondisi ekonomi, ketenagakerjaan, kualitas sumber daya manusia, serta karakteristik demografi yang secara keseluruhan turut memengaruhi tingkat pengangguran.

Perbedaan kondisi sosial ekonomi antar wilayah kabupaten/kota di Provinsi Kalimantan Barat menyebabkan karakteristik pengangguran di setiap wilayah tidak sama. Analisis pengangguran yang dilakukan secara umum belum mampu menunjukkan secara jelas persamaan dan perbedaan karakteristik tersebut sehingga berisiko menghasilkan kebijakan yang tidak relevan dengan kebutuhan spesifik daerah. Hal ini menunjukkan bahwa pengangguran tidak hanya menjadi permasalahan ketenagakerjaan, tetapi juga mencerminkan ketimpangan sosial dan ekonomi antar wilayah. Oleh karena itu, diperlukan kajian yang mampu mengidentifikasi pola dan kesamaan karakteristik pengangguran antar kabupaten/kota agar perumusan kebijakan ketenagakerjaan dapat dilakukan secara lebih efektif dan tepat sasaran.

Pendekatan *data mining* melalui metode *clustering* dapat digunakan untuk mengelompokkan kabupaten/kota ke dalam kelompok-kelompok yang memiliki karakteristik serupa. *Clustering* merupakan sebuah metode analisis multivariat yang bertujuan mengelompokkan objek ke dalam suatu kelompok sedemikian sehingga objek yang berada di dalam satu kelompok memiliki tingkat kesamaan yang sangat mirip dibandingkan dengan objek yang berada di kelompok yang lain (Pratama, Sulistianingsih, dan Imro'ah, 2024). Salah satu metode *clustering* yang populer yaitu algoritma *K-Means Clustering*. *K-Means* termasuk ke dalam *partitioning clustering*, yaitu mempartisi data menjadi beberapa kelompok yang saling lepas dengan jumlah *cluster* yang sudah ditetapkan di awal. Pada proses pembentukan *cluster*, *K-Means* memiliki kelemahan utama, yaitu keakuratan hasil sangat

dipengaruhi oleh pemilihan titik awal pusat *cluster*, sehingga berpotensi menghasilkan pembagian *cluster* yang kurang optimal (Sary, Satyahadewi, dan Andani, 2024).

Sebagai pengembangan dari *K-Means*, metode *K-Harmonic Means* diperkenalkan untuk mengurangi ketergantungan terhadap penentuan pusat *cluster* awal. Metode ini menggunakan fungsi objektif berbasis rata-rata harmonik jarak seluruh data terhadap pusat *cluster*, sehingga setiap titik data memiliki bobot yang bersifat dinamis (Christienova, Suparman, dan Zulhanif, 2017). Pendekatan ini memungkinkan proses pembentukan *cluster* berlangsung lebih stabil, khususnya pada data dengan karakteristik sosial ekonomi yang beragam (Yunistya, Goejantoro, dan Amijaya, 2022).

Selain proses pembentukan *cluster*, evaluasi terhadap kualitas hasil pengelompokan merupakan aspek penting yang perlu diperhatikan agar struktur *cluster* yang terbentuk memiliki tingkat kekompakan dan pemisahan antar *cluster* yang memadai. Oleh karena itu, penelitian ini menggunakan *Dunn Index* sebagai metode validasi internal untuk mengevaluasi kualitas hasil pengelompokan. *Dunn Index* mengukur rasio antara jarak minimum antar *cluster* dan jarak maksimum di dalam *cluster*, sehingga nilai *Dunn Index* yang lebih tinggi menunjukkan kualitas pengelompokan yang lebih baik (Malikhatin, Rusgiyono, dan Maruddani, 2021).

Integrasi antara proses pembentukan *cluster* menggunakan *K-Harmonic Means* dan evaluasi kualitas pengelompokan menggunakan *Dunn Index* diarahkan untuk memperoleh struktur pengelompokan wilayah kabupaten/kota yang tidak hanya terbentuk secara matematis, tetapi juga memiliki kualitas pengelompokan yang baik dan terukur secara kuantitatif. Melalui pendekatan tersebut, penelitian ini diharapkan mampu memberikan gambaran yang lebih komprehensif mengenai pola pengangguran di Provinsi Kalimantan Barat serta menjadi dasar pertimbangan dalam perumusan kebijakan ketenagakerjaan yang lebih efektif dan tepat sasaran.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan metode *K-Harmonic Means Clustering* dalam mengelompokkan kabupaten/kota di Provinsi Kalimantan Barat berdasarkan penyebab tingkat pengangguran?
2. Bagaimana menentukan jumlah *cluster* optimal menggunakan *Dunn Index* pada hasil pengelompokan kabupaten/kota di Provinsi Kalimantan Barat dengan metode *K-Harmonic Means*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah tersebut, tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan metode *K-Harmonic Means Clustering* dalam mengelompokkan kabupaten/kota di Provinsi Kalimantan Barat berdasarkan penyebab tingkat pengangguran.
2. Menentukan jumlah *cluster* optimal berdasarkan *Dunn Index* pada hasil pengelompokan kabupaten/kota di Provinsi Kalimantan Barat menggunakan metode *K-Harmonic Means*.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini difokuskan pada data numerik yang terkait dengan tingkat pengangguran di Provinsi Kalimantan Barat tahun 2024.
2. Jumlah *cluster* (K) pada penelitian ini ditentukan sebanyak 2,3, dan 4.
3. Parameter (p) pada penelitian ini ditentukan sebanyak 2; 2,5; 3; 3,5; dan 4.
4. Penelitian ini menggunakan pengukuran jarak *Manhattan* untuk menghitung jarak antar data dalam proses *clustering*.
5. Validasi hasil pengelompokan menggunakan *Dunn Index* dalam menilai kualitas *cluster* yang terbentuk.

1.5 Tinjauan Pustaka

Penelitian yang dilakukan oleh Shalsabila, Amalita, Vionanda, dan Permana (2023) menerapkan metode *K-Harmonic Means* (KHM) untuk mengelompokkan

kabupaten/kota di Indonesia berdasarkan tingkat kemiskinan. Penelitian ini dilatarbelakangi oleh kondisi tingkat kemiskinan di Indonesia yang meskipun secara nasional cenderung menurun, namun masih terdapat beberapa daerah yang mengalami peningkatan kemiskinan. Oleh karena itu, upaya penanggulangan kemiskinan tidak dapat dilakukan secara seragam dan perlu memperhatikan karakteristik masing-masing wilayah. Data yang digunakan merupakan data sekunder yang bersumber dari publikasi Badan Pusat Statistik (BPS) tahun 2022, dengan indikator meliputi persentase penduduk miskin, indeks kedalaman kemiskinan, indeks keparahan kemiskinan, dan garis kemiskinan. Proses analisis dilakukan melalui standarisasi data, pembentukan *cluster*, dan validasi hasil *clustering*. Hasil penelitian menunjukkan bahwa jumlah *cluster* optimal adalah dua kelompok, yaitu *cluster* pertama yang terdiri dari 54 kabupaten/kota dan *cluster* kedua yang mencakup 460 kabupaten/kota. Validasi menggunakan *Dunn Index* menghasilkan nilai sebesar 0,03492, yang menunjukkan bahwa metode *K-Harmonic Means* dengan parameter $p = 2{,}25$ memberikan hasil pengelompokan yang paling optimal.

Penelitian selanjutnya yaitu penelitian yang dilakukan oleh Sofia (2023) yang berisi tentang pengelompokan kabupaten/kota di Indonesia berdasarkan indikator Indeks Pembangunan Manusia (IPM) tahun 2022 menggunakan metode *K-Harmonic Means Clustering*. Penelitian ini bertujuan untuk mengidentifikasi karakteristik daerah berdasarkan variabel Umur Harapan Hidup (UHH), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Pengeluaran Riil per Kapita yang disesuaikan. Hasil *clustering* menunjukkan bahwa terdapat empat *cluster* utama: *cluster* dengan kategori sangat tinggi yang beranggotakan 316 kabupaten/kota, kategori tinggi dengan 182 kabupaten/kota, kategori sedang dengan 15 kabupaten/kota, dan kategori rendah dengan hanya satu kabupaten/kota. Dari 13 kali iterasi, metode ini menunjukkan tingkat kestabilan yang tinggi dengan akurasi 99,4%, dimana 511 kabupaten/kota tetap berada dalam *cluster* yang sama di setiap iterasi, sementara 3 kabupaten/kota mengalami perubahan *cluster*. Penelitian ini mengindikasikan adanya kesenjangan dalam indikator IPM antar wilayah, yang dapat menjadi dasar bagi kebijakan pemerataan pembangunan.

Yunistya, Goejantoro, dan Amijaya (2022) menerapkan metode *K-Harmonic Means (KHM)* untuk mengelompokkan kabupaten/kota di Pulau Kalimantan berdasarkan tingkat kemiskinan. Studi ini menggunakan data sekunder dengan mempertimbangkan beberapa indikator kemiskinan sebagai dasar pengelompokan. Metode *K-Harmonic Means* dipilih karena kemampuannya dalam mengatasi kelemahan *K-Means* terkait sensitivitas terhadap pemilihan pusat *cluster* awal, serta kemampuannya dalam menghasilkan *cluster* dengan persebaran data yang lebih kecil. Hasil penelitian menunjukkan bahwa jumlah *cluster* yang paling optimal adalah dua kelompok, dengan *cluster* pertama terdiri dari 11 kabupaten/kota dan *cluster* kedua mencakup 45 kabupaten/kota. Untuk memastikan validitas hasil pengelompokan, penelitian ini menggunakan *Silhouette Coefficient* sebagai alat evaluasi, yang menunjukkan bahwa metode *K-Harmonic Means* dengan parameter $p = 4$ menghasilkan nilai 0,323, yang mengindikasikan bahwa struktur *cluster* yang terbentuk masih lemah.

Immani, Sulistianingsih, dan Perdana (2024) menerapkan metode *K-Harmonic Means* untuk mengelompokkan kabupaten/kota di Provinsi Kalimantan Barat berdasarkan faktor-faktor yang memengaruhi *stunting*. Variabel yang dianalisis meliputi persentase rumah tangga tanpa akses air minum bersih, angka pemberian ASI eksklusif, persentase bayi berat lahir rendah yang lahir selamat, serta persentase rumah tangga tanpa sarana sanitasi layak. Evaluasi hasil pengelompokan dilakukan menggunakan *Silhouette Coefficient*, yang menunjukkan bahwa pembentukan *cluster* dengan parameter $p = 2$ memberikan hasil paling optimal, ditunjukkan dengan nilai koefisien 0,744 yang menandakan struktur *cluster* yang kuat.

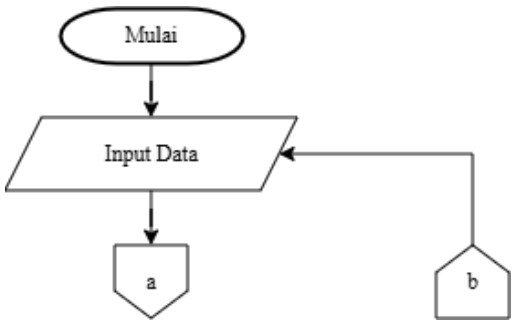
1.6 Metodologi Penelitian

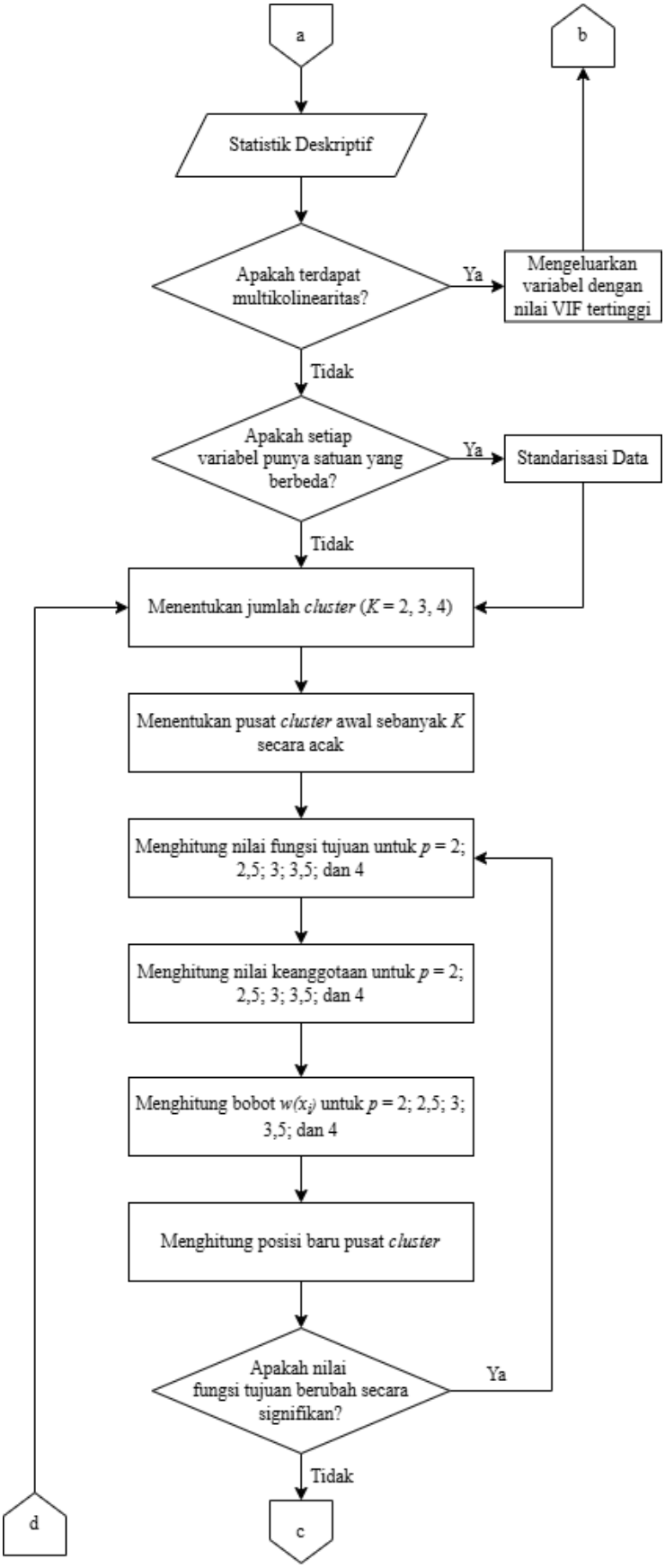
Proses penelitian dimulai dengan melakukan studi literatur yang relevan dengan topik penelitian serta pengumpulan data yang mencakup faktor-faktor penyebab pengangguran di Provinsi Kalimantan Barat. Setelah data diperoleh, dilakukan input data yang dilanjutkan dengan analisis statistika deskriptif untuk memahami karakteristik setiap variabel. Selanjutnya, dilakukan uji multikolinearitas untuk

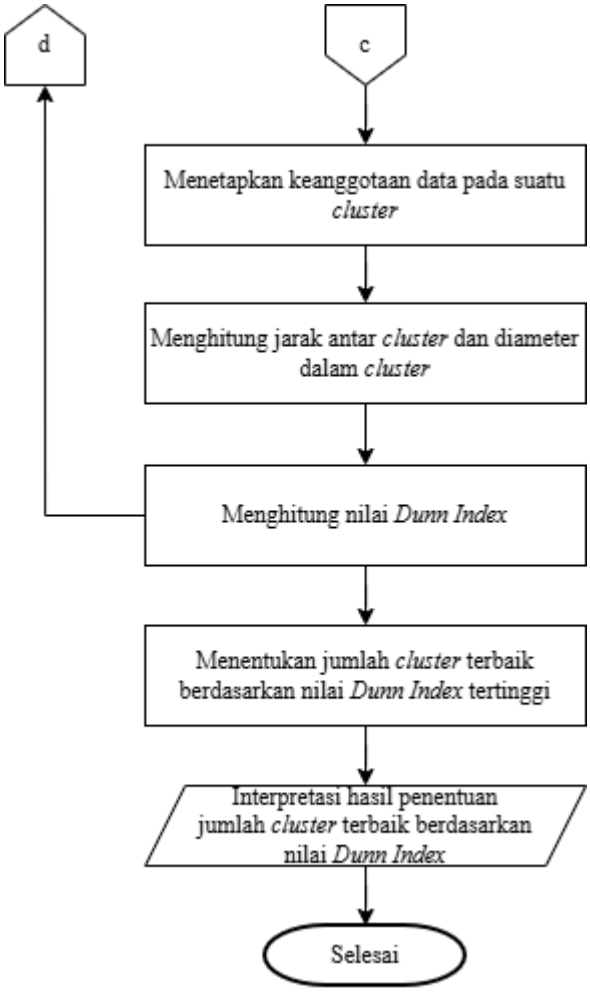
memastikan tidak terdapat korelasi yang sangat kuat antar variabel. Setelah itu, dilakukan standarisasi data menggunakan *Z-score* untuk menyamakan skala dan satuan antar variabel agar hasil perhitungan jarak dalam algoritma *clustering* menjadi objektif.

Tahap selanjutnya adalah proses *clustering* menggunakan metode *K-Harmonic Means*. Proses ini diawali dengan penentuan jumlah *cluster*, yaitu $K = 2, 3,$ dan 4. Setelah jumlah *cluster* ditetapkan, dilakukan inisialisasi pusat *cluster* awal secara acak. Proses iterasi kemudian dilakukan dengan menghitung nilai keanggotaan setiap data terhadap *cluster* untuk variasi parameter $p = 2;\ 2{,}5;\ 3;\ 3{,}5;$ dan 4, dilanjutkan dengan perhitungan bobot data serta pembaruan posisi pusat *cluster*. Evaluasi fungsi tujuan dilakukan pada setiap iterasi untuk mengetahui apakah terjadi perubahan yang signifikan. Iterasi akan dihentikan apabila nilai fungsi tujuan telah konvergen dan posisi pusat *cluster* tidak mengalami perubahan yang signifikan, selanjutnya ditetapkan keanggotaan data pada masing-masing *cluster*.

Setelah proses *clustering* selesai, dilakukan validasi hasil pengelompokan menggunakan *Dunn Index*. Pada tahap ini dihitung jarak intra-*cluster* dan jarak inter-*cluster* untuk memperoleh nilai *Dunn Index* pada setiap jumlah *cluster*. Jumlah *cluster* terbaik ditentukan berdasarkan nilai *Dunn Index* tertinggi. Tahap akhir penelitian adalah melakukan interpretasi hasil penentuan jumlah *cluster* terbaik berdasarkan nilai *Dunn Index*, sehingga diperoleh gambaran pengelompokan tingkat pengangguran di Provinsi Kalimantan Barat sesuai dengan karakteristik masing-masing *cluster*. Adapun tahapan yang diimplementasikan divisualisasikan pada *flowchart* penelitian berikut.




a
b
Statistik Deskriptif
Apakah terdapat multikolinearitas?
Ya
Mengeluarkan variabel dengan nilai VIF tertinggi
Tidak
Apakah setiap variabel punya satuan yang berbeda?
Ya
Standarisasi Data
Tidak
Menentukan jumlah *cluster* (K = 2, 3, 4)
Menentukan pusat *cluster* awal sebanyak K secara acak
Menghitung nilai fungsi tujuan untuk p = 2; 2,5; 3; 3,5; dan 4
Menghitung nilai keanggotaan untuk p = 2; 2,5; 3; 3,5; dan 4
Menghitung bobot $w(x_j)$ untuk p = 2; 2,5; 3; 3,5; dan 4
Menghitung posisi baru pusat *cluster*
Apakah nilai fungsi tujuan berubah secara signifikan?
Ya
Tidak
d
c



Gambar 1.1 *Flowchart* Penelitian