PENERAPAN *K-MEDOIDS CLUSTERING* MENGGUNAKAN VALIDASI *C-INDEX* DALAM PENGELOMPOKAN DAERAH DI KALIMANTAN BARAT BERDASARKAN INDIKATOR KETAHANAN PANGAN

KHALISHAH GHINA FALLAH
NIM H1091211068

SKRIPSI

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

PENERAPAN *K-MEDOIDS CLUSTERING* MENGGUNAKAN VALIDASI *C-INDEX* DALAM PENGELOMPOKAN DAERAH DI KALIMANTAN BARAT BERDASARKAN INDIKATOR KETAHANAN PANGAN

KHALISHAH GHINA FALLAH
NIM H1091211068

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	: Penerapan *K-Medoids Clustering* Menggunakan Validasi *C-Index* Dalam Pengelompokan Daerah di Kalimantan Barat Berdasarkan Indikator Ketahanan Pangan
Nama Mahasiswa	: Khalishah Ghina Fallah
NIM	: H1091211068
Jurusan/Program Studi	: Matematika/Statistika
Tanggal Lulus	: 18 Juni 2025
SK Pembimbing	: No. 2641/UN22.8/TD.06/2024/Tanggal 2 September 2024
SK Penguji	: No. 1733/UN22.8/TD.06/2025/Tanggal 10 Juni 2025

Dosen Pembimbing

Pembimbing I	Pembimbing II
Shantika Martha, M.Si	Hendra Perdana, S.Si., M.Sc.
NIP 19840308200812200 3	NIP 198810102019031020

Dosen Penguji

Ketua Penguji	Anggota Penguji
Dr. Evy Sulistianingsih, S.Si., M.Sc	Nurfitri Imro'ah, M.Si
NIP 198502172008122006	NIP 198907182019032021

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Shantika Martha, M.Si	Hendra Perdana, S.Si., M.Sc.
NIP 19840308200812200 3	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	: Khalishah Ghina Fallah
NIM	: H1091211068
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, 18 Juni 2025



Khalishah Ghina Fallah

NIM. H1091211068

Penerapan *K-Medoids Clustering* Menggunakan Validasi *C-Index* Dalam Pengelompokan Daerah di Kalimantan Barat Berdasarkan Indikator Ketahanan Pangan

Abstrak

Ketahanan pangan merupakan kondisi terpenuhinya kebutuhan pangan bagi setiap individu dalam suatu wilayah, baik dari segi ketersediaan, akses, maupun pemanfaatannya, sehingga masyarakat dapat hidup sehat dan produktif. Analisis ketahanan pangan diperlukan untuk memahami karakteristik wilayah berdasarkan berbagai indikator yang berpengaruh. Untuk mengukur ketahanan pangan membutuhkan ukuran yang komprehensif dengan melibatkan serangkaian indikator yakni aspek ketersediaan, akses dan pemanfaatan pangan. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Kalimantan Barat berdasarkan indikator ketahanan pangan serta menentukan jumlah kelompok atau *cluster* optimal yang terbentuk sehingga karakteristik setiap kelompok dapat diketahui melalui analisis *cluster*. Metode *K-Medoids* merupakan salah satu algoritma yang efektif dalam pengelompokan karena mampu menghasilkan *cluster* yang lebih stabil dan representatif terhadap data yang digunakan. Namun, penentuan jumlah *cluster* optimal masih belum memiliki dasar teori yang kuat, sehingga diperlukan metode untuk menentukannya. *C-Index* adalah metode validasi internal yang menilai kualitas *clustering* dengan membandingkan jarak antar objek dalam dan antar *cluster*. Berdasarkan penelitian yang telah dilakukan, jumlah *cluster* optimal yang terbentuk berjumlah empat *cluster* dengan nilai *C-index* terkecil sebesar 0,42. *Cluster* 1 merupakan daerah dengan ketahanan pangan yang tinggi yang terdiri dari 5 kabupaten/kota, *cluster* 2 dengan tingkat ketahanan pangan yang cukup terdiri dari empat kabupaten/kota, *cluster* 3 dengan tingkat ketahanan pangan yang kurang terdiri dari empat kabupaten/kota, dan *cluster* 4 dengan tingkat ketahanan pangan yang rendah yang terdiri dari satu kabupaten/kota yaitu Kayong Utara.

Kata kunci: Standarisasi, Jarak *Euclidean, Cluster* Optimal

Application of K-Medoids Clustering Using C-Index Validation in Grouping Regions in West Kalimantan Based on Food Security Indicators

Abstract

Food security is a condition where the food needs of every individual in a region are met, both in terms of availability, access, and utilization, so that people can live healthy and productive lives. Food security analysis is needed to understand the characteristics of a region based on various influential indicators. To measure food security, a comprehensive measure is needed involving a series of indicators, namely aspects of food availability, access and utilization. This study aims to group districts/cities in West Kalimantan based on food security indicators and determine the number of optimal groups or clusters formed so that the characteristics of each group can be known through cluster analysis. The K-Medoids method is one of the effective algorithms in clustering because it is able to produce clusters that are more stable and representative of the data used. However, determining the optimal number of clusters still does not have a strong theoretical basis, so a method is needed to determine it. C-Index is an internal validation method that assesses the quality of clustering by comparing the distance between objects within and between clusters. Based on the research that has been conducted, the number of optimal clusters formed is four clusters with the smallest C-index value of 0,42. Cluster 1 is an area with high food security consisting of 5 districts/cities, cluster 2 with a sufficient level of food security consisting of four districts/cities, cluster 3 with a low level of food security consisting of four districts/cities, and cluster 4 with a low level of food security consisting of one district/city, namely Kayong Utara.

Keywords: Standardization, Euclidean Distance, Optimal Cluster

PRAKATA

Dengan mengucap puji dan syukur penulis panjatkan kepada Allah Subhanahu wata'ala atas segala rahmat dan hidayah-Nya, serta perlindungan, pertolongan, dan ridho-Nya, tidak terlupakan iringan salam dan sholawat bagi junjungan kita Nabi Muhammad Sallahu Alaihi Wassalam, sehingga penulis dapat menyelesaikan Tugas Akhir yang berujudul "**Penerapan *K-Medoids Clustering* Menggunakan Validasi *C-Index* Dalam Pengelompokan Daerah di Kalimantan Barat Berdasarkan Indikator Ketahanan Pangan**" dengan tepat pada waktunya. Penulisan skripsi ini bertujuan untuk memenuhi syarat memperoleh gelar Sarjana Statistika pada Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Tanjungpura.

Dalam penulisan skripsi ini tidak luput dari bimbingan, dukungan, serta motivasi dan doa dari berbagai pihak, sehingga penulis dapat menyelesaikan skripsi dengan lancar. Penulis menyampaikan penghargaan dan terima kasih kepada:

1. Allah SWT yang telah melimpahan rahmat dan berkahnya serta memberikan kelancaran sehingga penulis dapat menyelesaikan penulisan skripsi ini dapat di selesaikan.
2. Keluarga tercinta, terima kasih telah mengusahakan segalanya untuk anak bungsumu yang satu ini. Terima kasih atas doa hebat yang selalu kalian panjatkan untuk menulis, semoga kalian sehat selalu dan selalu ada dalam setiap episode kehidupan penulis. Penulis meminta maaf jika belum bisa memberikan yang terbaik dan penulis berharap suatu saat nanti kalian bangga dengan anak perempuan satu-satu nya ini.
3. Ibu Shantika Martha, M.Si. selaku dosen pembimbing pertama yang telah meluangkan waktu, mengarahkan, memotivasi, masukkan, saran, dan membimbing selama penulisan skripsi ini.
4. Bapak Hendra Perdana, M.Sc.. selaku dosen pembimbing pertama yang telah meluangkan waktu, mengarahkan, memotivasi, masukkan, saran, dan membimbing selama penulisan skripsi ini.
5. Dr. Evy Sulistianingsih, M.Sc. selaku dosen penguji pertama yang telah memberikan saran, masukan, serta koreksi dalam penulisan skripsi ini.

6. Nurfitri Imro'ah, M.Si. selaku dosen penguji kedua yang telah memberikan saran, masukan, serta koreksi dalam penulisan skripsi ini.
7. Seluruh Bapak/Ibu Dosen Jurusan Matematika/Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam yang telah memberikan ilmu bermanfaat selama proses perkuliahan.
8. Saudara Egi Kusnadi. terima kasih selalu menemani dan selalu menjadi support system penulis pada hari yang tidak mudah selama proses pengerjaan skripsi ini. Terima kasih telah mendengar keluh kesah, memberikan dukungan, semangat, materi, maupun menemani di waktu libur kerjamu. Terima kasih telah menjadi bagian perjalanan saya hingga penyusunan skripsi ini.
9. Sahabat-sahabat penulis yaitu Wahida, Nanda, Hera, Della, Putri, Luis, Muti, Siska, dan Ulya yang telah membersamai dan saling menyemangati selama perkuliahan hingga saat proses penulisan skripsi ini selesai.
10. Teman-teman "Expos" Statistika Angkatan 2021 yang membersamai penulis selama proses kuliah di Program Studi Statistika FMIPA Untan.
11. Seluruh pihak yang tidak dapat disebutkan satu persatu yang telah memberikan doa dan semangat kepada penulis dalam penyelesaian skripsi ini.

Semoga segala hal yang telah diberikan kepada penulis menjadi amal ibadah serta diberkahi Allah Subhanahu wata'ala. Penulis menyadari penyusunan skripsi ini jauh dari kata sempurna maka dari itu penulis menerima saran dan kritikan yang membangun.

Pontianak, 18 Juni 2025



Khalishah Ghina Fallah

DAFTAR ISI

LEMBAR PENGESAHAN ii
PERNYATAAN INTEGRITAS AKADEMIK iii
Abstrak iv
Abstract v
PRAKATA vi
DAFTAR ISI viii
DAFTAR GAMBAR x
DAFTAR TABEL xi
DAFTAR LAMPIRAN xii
DAFTAR SIMBOL xiii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 4
1.4 Batasan Masalah 4
1.5 Tinjauan Pustaka 5
1.6 Metodologi Penelitian 6
BAB II LANDASAN TEORI 10
2.1 Ketahanan Pangan 10
2.2 Analisis Multivariat 11
2.3 Metode non-Hirarki 12
2.4 Multikolinearitas 13
2.5 Standarisasi 14
2.6 *Euclidean Distance* 15
BAB III *K-MEDOIDS* DAN *C-INDEX* 17
3.1 Pengertian Algoritma *K-medoids* 17
3.2 *C-index* 19
BAB IV HASIL DAN PEMBAHASAN 21

4.1 Deskripsi Data 21
4.2 Pengujian Multikolinearitas 22
4.3 Standarisasi Data 23
4.4 *Cluster K-medoids* 23
4.5 Penentuan Banyak *Cluster* Menggunakan *C-index* 28
4.6 Interpretasi Profil *Cluster* 33
BAB V KESIMPULAN 37
DAFTAR PUSTAKA 38
LAMPIRAN 42
Lampiran 1 Data Penelitian 42
Lampiran 2 Standarisasi Data 43
Lampiran 3 Jarak *Euclidean* 44

DAFTAR GAMBAR

Gambar 1.1 Flowchart Penelitian 9
Gambar 4.1 Peta Hasil Pengelompokan 32

DAFTAR TABEL

Tabel 4.1 Statistik Deskriptif 21
Tabel 4.2 Nilai *Variance Inflation Factor* (VIF) 23
Tabel 4.3 Pembagian *Cluster* Berdasarkan Medoids Awal 26
Tabel 4.4 Pembagian *Cluster* Berdasarkan Medoids Baru 27
Tabel 4.5 Nilai *Sw* 29
Tabel 4.6 Nilai *C-index* pada Setiap *Cluster* 31
Tabel 4.7 Pengelompokan Kab/Kota di Kalimantan Barat pada 4 *Cluster* 33
Tabel 4.8 Karakteristik Tiap *Cluster* 33

DAFTAR LAMPIRAN

Lampiran 1 Data Penelitian 40
Lampiran 2 Standarisasi Data 41
Lampiran 3 Jarak *euclidean* 42

DAFTAR SIMBOL

$\boldsymbol{p}$	:	Banyaknya variabel yang diamati
$\boldsymbol{n}$	:	Banyaknya objek/data yang digunakan
$\boldsymbol{K}$	:	Banyaknya *cluster* yang digunakan
$\boldsymbol{R_j^2}$	:	Nilai koefisien determinasi variabel dependen dan independen.
$\boldsymbol{X_j}$	:	Variabel ke-*j* dengan $j = 1,2,\ldots,p$
$\boldsymbol{x_{i,j}}$	:	Nilai objek ke-*i* untuk variabel ke-*j*
$\boldsymbol{\mu_j}$	:	Rata-rata variabel ke-*j*
$\boldsymbol{\sigma_j}$	:	Nilai standar deviasi ke-*j*
$\boldsymbol{Z_{i,j}}$	:	Nilai standarisasi objek ke-*i* pada variabel ke-*j* dengan $i = 1,2,\ldots,n$
$\boldsymbol{d_{ic}}$	:	Jarak *euclidean* dari objek ke-i ke pusat *cluster* ke-*c*
$\boldsymbol{x_{ik}}$	:	Nilai objek ke-*i* pada variabel ke-*k*, dengan $i=1,2,\ldots,n$ dan $k=1,2,\ldots,p$
$\boldsymbol{y_{ck}}$	:	Nilai pusat *cluster* ke-*c* pada variabel ke-*k*, dengan $c=1,2,\ldots,K$ dan $k=1,2,\ldots,p$
$\boldsymbol{S_w}$	:	Jumlah jarak antar objek dalam satu *cluster*
$\boldsymbol{S_{min}}$	:	Jumlah jarak terkecil antara semua pasangan objek data dalam satu *cluster* dan antar *cluster*
$\boldsymbol{S_{max}}$	:	Jumlah jarak terbesar antara semua pasangan objek data dalam satu *cluster* dan antar *cluster*
$\boldsymbol{m}$	:	Jumlah pasangan elemen dalam *cluster*
$\boldsymbol{b}$	:	Indeks *cluster* dengan $b=1,2,\ldots,K$.
$\boldsymbol{A_b}$	:	Jumlah anggota dalam *cluster* ke-*b*.

BAB I
PENDAHULUAN

1.1 Latar Belakang

Indonesia pernah menghadapi krisis ekonomi pada tahun 1998, yang menguji ketahanan ekonomi serta berdampak luas terhadap berbagai sektor, termasuk ketahanan pangan. Krisis tersebut menyebabkan lonjakan harga bahan pokok, penurunan daya beli masyarakat, meningkatnya angka kemiskinan, serta terbatasnya akses terhadap pangan yang bergizi dan berkualitas. Meskipun pemulihan yang dilakukan berhasil mendorong pertumbuhan ekonomi yang relatif stabil, kesenjangan dalam ketahanan pangan masih menjadi tantangan di berbagai daerah, terutama di daerah yang memiliki tingkat kerentanan sosial dan ekonomi yang tinggi. Hal ini menunjukkan bahwa meskipun terdapat kemajuan di sektor makro ekonomi, berbagai masalah dalam kehidupan masyarakat, seperti rendahnya akses terhadap pendidikan, kesehatan, dan kebutuhan dasar lainnya, masih menjadi tantangan utama dalam mencapai kesejahteraan yang merata bagi seluruh rakyat Indonesia. Sebagai negara yang menandatangani World Food Summit (1996) dan Millennium Declaration (2000), Indonesia terus memperkuat upayanya untuk mencapai tujuan utama Millennium Development Goals (MDG), yaitu mengurangi jumlah penduduk yang hidup dengan pendapatan di bawah US$ 1 per hari serta menurunkan angka kelaparan menjadi setengahnya pada tahun 2015 (Mulyasari, 2016)

Pangan adalah kebutuhan dasar suatu negara atau perseorangan dari segi ragam, jumlah, mutu, keamanan, dan gizi, dimana kondisi terpenuhnya kebutuhan ini dikenal dengan istilah ketahanan pangan (Wulandari dan Fauzy, 2016). Di Indonesia, Undang-Undang No. 7 tahun 1996 tentang pangan mengartikan ketahanan pangan sebagai kondisi terpenuhinya pangan bagi rumah tangga yang tercermin dari tersedianya pangan yang cukup, baik jumlah maupun mutunya, merata, terjangkau dan aman (Yamin, Putri, dan Mulyana, 2019). Ketahanan pangan adalah kondisi di mana pangan tersedia secara cukup, aman,

bergizi, merata, dan terjangkau bagi semua orang, sesuai dengan agama, keyakinan, dan budaya, sehingga mereka dapat hidup sehat, aktif, dan produktif secara berkelanjutan.

Ketahanan pangan menjadi isu penting seiring perkembangan zaman. Di Indonesia, ketahanan pangan menjadi perhatian utama pemerintah dan masyarakat karena pangan bukan hanya kebutuhan dasar, tetapi juga bagian dari hak asasi manusia yang harus dipenuhi. Hal ini diatur secara tegas dalam Undang-Undang No. 18 Tahun 2012 tentang Pangan, yang menegaskan bahwa penyediaan pangan yang cukup, aman, dan bergizi adalah kewajiban negara untuk menjamin kesehatan dan kualitas hidup warganya. Dengan terpenuhinya ketahanan pangan, diharapkan sumber daya manusia Indonesia dapat tumbuh dan berkembang secara optimal, sehingga mendukung pembangunan nasional yang berkelanjutan (Peraturan Pemerintah, 2019). Perubahan lingkungan seperti perubahan iklim, degradasi lahan, dan bencana alam mengancam produksi dan distribusi pangan, yang dapat mengganggu ketahanan pangan. Hal ini menjadi tantangan besar, terutama dengan pertumbuhan jumlah penduduk Indonesia. Jika ketahanan pangan tidak dikelola dengan baik, risiko kelaparan dan kemiskinan akan meningkat, sehingga mempengaruhi kualitas hidup dan menghambat pembangunan berkelanjutan. Upaya meningkatkan ketahanan pangan sangat penting, terutama di negara berkembang, karena memiliki dua peran utama. Pertama, memastikan ketersediaan pangan yang cukup, sehat, dan berkualitas bagi masyarakat. Kedua, mendukung pengembangan kreativitas, inovasi, serta produktivitas di berbagai bidang, termasuk ilmu pengetahuan, teknologi, dan ketenagakerjaan (Dharshinni dan Fandi, 2022).

Populasi Kalimantan Barat pada tahun 2021 mencapai 5,47 juta jiwa, dengan 17,15% di antaranya tinggal di perkotaan. Urbanisasi yang berkembang mengubah pola konsumsi masyarakat, termasuk meningkatnya preferensi terhadap pangan bergizi. Meskipun memiliki sumber daya alam yang melimpah, ketahanan pangan masih menghadapi tantangan akibat keterbatasan infrastruktur, distribusi pangan yang tidak merata, dan rendahnya akses terhadap pangan bergizi. Berdasarkan hasil analisis Peta Ketahanan dan Kerentanan

Pangan Kalimantan Barat tahun 2022, jumlah kecamatan yang tergolong rentan pangan mengalami penurunan dibandingkan tahun 2021, dari 33 kecamatan (18,98%) menjadi 31 kecamatan (17,82%) dari total 174 kecamatan. Meskipun angka ini menunjukkan perbaikan, kondisi ketahanan pangan di Kalimantan Barat masih perlu mendapat perhatian serius. Beberapa kecamatan yang sebelumnya tidak tergolong rentan justru mengalami peningkatan risiko, sementara perubahan distribusi wilayah rentan menunjukkan adanya dinamika yang kompleks dalam ketahanan pangan. Faktor-faktor seperti urbanisasi, perubahan konsumsi, keterbatasan akses pangan, dan dampak ekonomi global dapat mengancam stabilitas pangan di masa depan. Oleh karena itu, analisis yang lebih mendalam sangat diperlukan untuk mengidentifikasi pola ketahanan pangan di setiap wilayah khususnya di Kalimantan Barat.

Untuk mengukur ketahanan pangan membutuhkan ukuran yang komprehensif dengan melibatkan serangkaian indikator yakni aspek ketersediaan, akses dan pemanfaatan pangan. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Kalimantan Barat berdasarkan indikator ketahanan pangan menggunakan analisis *cluster*. Analisis *cluster* adalah teknik statistik yang bertujuan untuk menggelompokkan suatu objek ke dalam sebuah kelompok sehingga objek dalam satu kelompok menunjukkan kemiripan yang tinggi terhadap objek dalam kelompok lain (Silvi, 2018). Salah satu metode analisis *cluster* yang dapat digunakan yaitu metode non hierarki

Metode non hirarki yang paling terkenal dan umum digunakan adalah *k-means dan k-medoids clustering*. *K-means clustering* adalah salah satu metode *partitional clustering* yang berguna untuk mengolah dan mengelompokkan data menjadi satu atau lebih *cluster*. *K-means* akan berjalan jika jumlah *cluster* diketahui terlebih dahulu, kemudian sebanyak *K* titik dipilih secara acak sebagai *centroid* (pusat *cluster*) (Syahraini & Rizal, 2024). Namun, *k-means* memiliki kelemahan yaitu sangat sensitif terhadap pencilan karena pusat *cluster* berupa rata-rata yang dapat terpengaruh oleh data ekstrem. Untuk mengatasi kelemahan tersebut, algoritma *k-medoids* digunakan karena menggunakan *medoid*s, yaitu data yang benar-benar ada dalam *cluster* sebagai pusatnya, sehingga lebih tahan

terhadap pencilan dan menghasilkan hasil *clustering* yang lebih stabil dan representatif (Rudianto & Wijayanto, 2024). Dalam analisis *cluster*, objek dikelompokkan berdasarkan kemiripan, yang diukur dengan jarak yaitu *euclidean distance*. Validasi hasil *cluster* penting untuk menentukan jumlah *cluster* yang tepat. Hal ini bertujuan agar hasil pengelompokan benar-benar mencerminkan struktur data yang sesungguhnya, sehingga informasi yang diperoleh dapat digunakan sebagai dasar pengambilan keputusan yang lebih akurat dan objektif. Oleh karena itu, dilakukan pengelompokan kabupaten/kota di Kalimantan Barat berdasarkan indikator-indikator yang digunakan dalam mengukur ketahanan pangan menggunakan *k-medoids* dan divalidasi dengan *c-index* sebagai metode validasi dilakukan karena mampu mengukur kualitas *cluster* berdasarkan distribusi jarak antar objek. Tujuan penelitian ini adalah memberikan rekomendasi berbasis data guna mendukung kebijakan ketahanan pangan di Kalimantan Barat.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka rumusan masalah dalam penelitian ini yaitu:

1. Bagaimana mengelompokkan Kabupaten/Kota berdasarkan indikator ketahanan pangan di Kalimantan Barat dengan algoritma *k-medoids?*
2. Bagaimana menentukan jumlah *cluster* optimal dalam mengelompokkan Kabupaten/Kota di Kalimantan Barat berdasarkan indikator ketahanan pangan dengan menggunakan validasi *c-index*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat maka tujuan penelitian yang ingin dicapai dalam penelitian ini yaitu:

1. Mengelompokkan Kabupaten/Kota di Kalimantan Barat berdasarkan indikator ketahanan pangan menggunakan algoritma *k-medoids.*
2. Menentukan jumlah cluster optimal pada pengelompokan Kabupaten/Kota di Kalimantan Barat berdasarkan indikator ketahanan pangan menggunakan validasi *c-index*.

1.4 Batasan Masalah

Berdasarkan uraian dari latar belakang, maka batasan masalah pada penelitian ini yaitu:

1. Pembentukan *cluster* optimal dibatasi maksimal sebanyak lima untuk menghindari *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri dengan data sehingga hasilnya tidak lagi mewakili pola umum yang sebenarnya. Pembatasan ini dilakukan untuk memastikan bahwa setiap *cluster* yang terbentuk memiliki perbedaan yang bermakna.
2. Pengukuran jarak yang digunakan adalah *euclidean distance*.
3. Variabel yang digunakan pada penelitian ini, yaitu rasio konsumsi normatif per kapita terhadap ketersediaan bersih; persentase penduduk yang hidup di bawah garis kemiskinan; persentase rumah tangga dengan proporsi pengeluaran untuk pangan lebih dari 65% terhadap total pengeluaran; persentase rumah tangga tanpa akses listrik; rata-rata lama sekolah perempuan di atas 15 tahun; persentase rumah tangga tanpa akses ke air layak; rasio jumlah penduduk per tenaga kesehatan terhadap kepadatan penduduk; persentase balita dengan tinggi badan dibawah standar; dan persentase angka kesakitan.

1.5 Tinjauan Pustaka

Penelitian yang dilakukan oleh Marlina, Putri, Fernando, dan Ramadhan (2018) bertujuan untuk mengevaluasi kinerja algoritma *k-medoids* dan *k-means* dalam pengelompokan data sebaran, khususnya pada data sebaran anak cacat. Hasil penelitian menunjukkan bahwa algoritma *k-medoids* memiliki kinerja yang lebih baik dalam pengelompokan data dibandingkan dengan *k-means*. Hal ini dibuktikan dengan validitas *silhouette coefficient* yang dihasilkan, di mana algoritma *k-medoids* memperoleh nilai sebesar 0,50, sedangkan *k-means* hanya mencapai 0,144. Dengan demikian, *k-medoids* lebih efektif dalam membentuk *cluster* yang lebih representatif terhadap pola sebaran anak cacat.

Nahdliyah, Widiharih, dan Prahutama (2019) telah melakukan penelitian dengan menggunakan *k-medoids* dalam mengelompokkan kriminalitas di

kabupaten/kota di Jawa Tengah. Pada penelitian ini, menggunakan *silhouette index* dan *c-index* untuk memilih jarak dan validasi terbaik. Penelitian ini menggunakan dua ukuran jarak yaitu jarak *euclidean* dan jarak *cityblock*. Hasil penelitian menunjukkan nilai *silhouette index* terbesar pada metode *k-medoids cluster* yaitu sebesar 0,389, sedangkan nilai *c-index* terkecil yaitu sebesar 0,043. Berdasarkan hasil penelitian, pengklasteran dengan dua validasi *silhouette index* dan *c-index* dapat disimpulkan bahwa hasil pengklasteran terbaik dengan kedua validasi tersebut berdasarkan data jumlah kriminalitas di Jawa Tengah tahun 2018 dengan metode *k-medoids* yaitu pengklasteran pada k= 4 dan pengukuran jarak *euclidean*, karena memiliki nilai indeks validasi yang paling optimum pada masing-masing validasi.

Saidah, Santoso, dan Widiharih (2022) melakukan penelitian dengan menggunakan *k-medoids* dalam mengelompokkan provinsi di Indonesia berdasarkan indikator kesehatan lingkungan. Pada penelitian ini, menggunakan *calinski-harabasz index, baker-hubert index, silhouette index, c-index,* dan *davies-bouldin index* untuk memilih jarak dan validasi terbaik. Penelitian ini menggunakan dua ukuran jarak yaitu jarak *euclidean* dan jarak *manhattan*. Hasil penelitian menunjukkan jumlah *cluster* terbaik adalah dua *cluster* dengan menggunakan metode pengukuran jarak *manhattan*, dengan nilai indeks *calinski-harabasz* terbesar = 24.10072, indeks *baker-hubert* terbesar = 0.8466251, indeks *silhouette* terbesar = 0.4246581, *c-index* terkecil = 0.07290109, dan *davies-bouldin index* terkecil = 1.094805.

1.6 Metodologi Penelitian

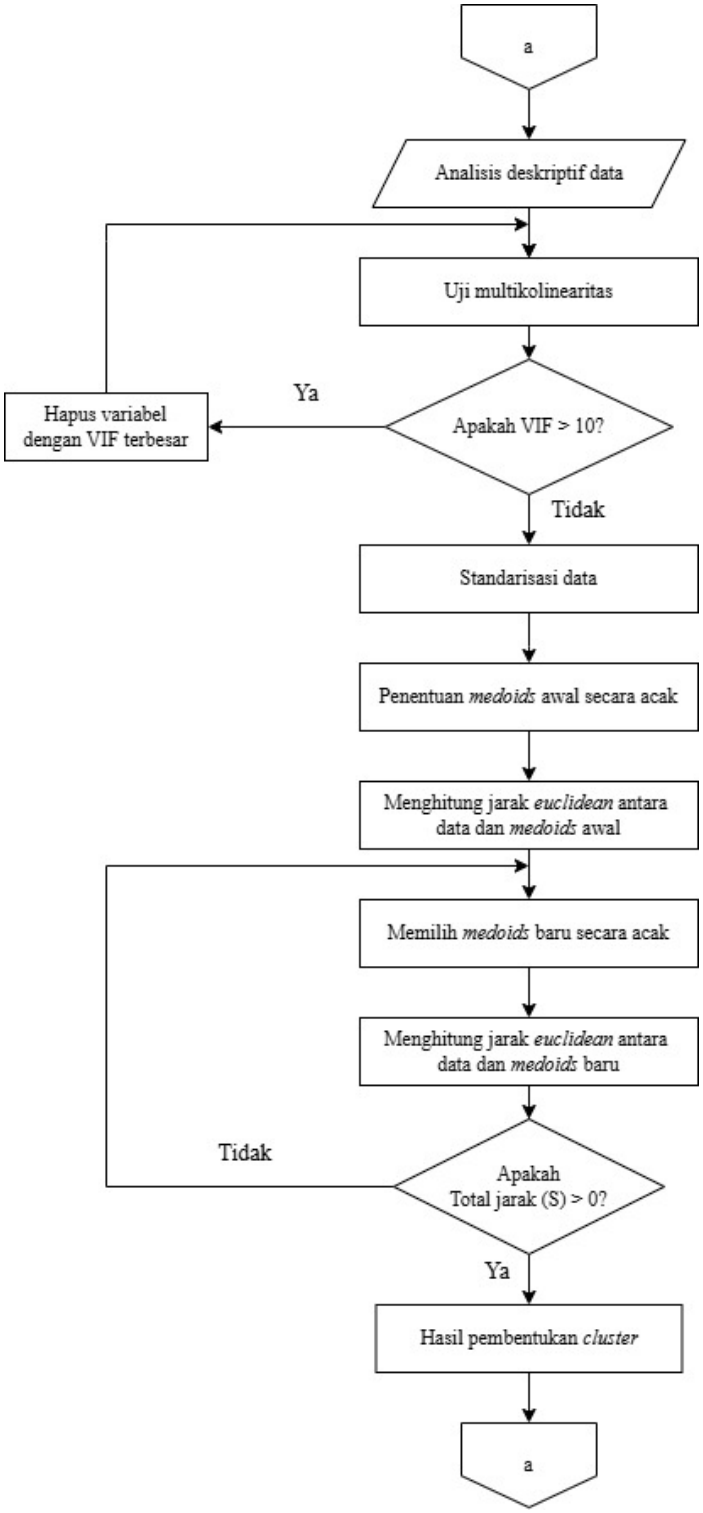
Data dalam penelitian ini diperoleh dari Badan Pangan Nasional melalui publikasi Peta Ketahanan dan Kerentanan Pangan Provinsi Kalimantan Barat Tahun 2023. Data yang digunakan mencakup indikator ketahanan pangan dari seluruh kabupaten/kota di Kalimantan Barat untuk tahun 2023. Metode yang digunakan yaitu metode *k-medoids* dengan optimasi *c-index* dengan jumlah data sebanyak 14 kabupaten/kota. Variabel penelitian ini yaitu Rasio konsumsi normatif terhadap produksi pangan (X_1), Persentase penduduk yang hidup dibawah garis kemiskinan (X_2), Persentase RT dengan proporsi pengeluaran pangan ≥ 65 persen terhadap

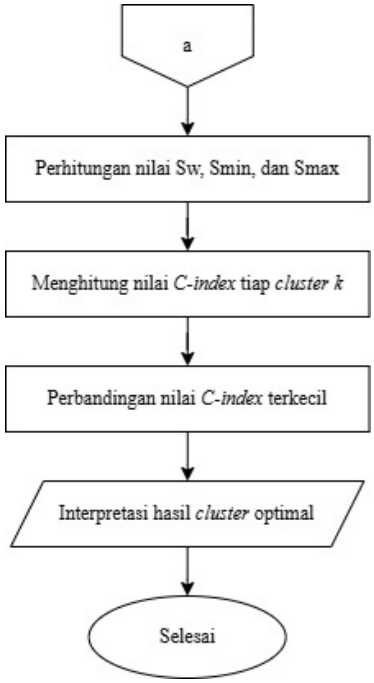
total pengeluaran (X_3), Persentase RT tidak memiliki akses listrik (X_4), Rata-rata lama sekolah perempuan usia 15 tahun ke atas (X_5), Persentase RT tanpa akses ke air layak (X_6), Rasio penduduk per tenaga kesehatan terhadap kepadatan penduduk (X_7), Presentase balita pendek dan sangat pendek (X_8), dan Persentase angka kesakitan (X_9).

Proses analisis dilakukan dengan bantuan software R*Studio* dan *Microsoft excel*. Langkah yang pertama dilakukan adalah analisis statistik deskriptif pada setiap variabel yang digunakan. Setelah itu melakukan proses standarisasi data. Kemudian melakukan pengujian multikolinearitas untuk melihat korelasi antar variabel-variabel dengan menghitung nilai *Variance Inflation Factor* (VIF). Jika nilai VIF pada semua variabel kurang dari sama dengan 10, maka tidak terjadi multikolinearitas sehingga proses penelitian dapat diteruskan ke langkah selanjutnya. Namun, jika nilai VIF pada satu atau beberapa variabel yang lebih dari 10, maka perlu dilakukan pembuangan variabel. Proses pengelompokan dimulai dengan menentukan jumlah *cluster* yang akan dibentuk, dimana dalam penelitian ini, digunakan 4 *cluster* sebagai contoh. Kemudian, memilih objek secara acak untuk digunakan sebagai *medoids* awal dan dilanjutkan dengan menghitung jarak *euclidean distance* dengan objek yang bukan *medoids* dimasukkan ke dalam kelompok dengan jarak terdekat terhadap pusat *cluster*. Setelah itu, memilih secara acak *medoids* baru dan menghitung *euclidean distance* pada *medoids* baru. Selanjutnya, menghitung total jarak (S) dan mengulangi kembali seluruh proses hingga iterasi berhenti apabila nilai total jarak (S) > 0. Setelah memperoleh hasil *clustering*, kemudian menentukan jumlah *cluster* optimal dengan *c-index*.

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




a
Analisis deskriptif data
Uji multikolinearitas
Apakah VIF > 10?
Ya
Hapus variabel dengan VIF terbesar
Tidak
Standarisasi data
Penentuan *medoids* awal secara acak
Menghitung jarak *euclidean* antara data dan *medoids* awal
Memilih *medoids* baru secara acak
Menghitung jarak *euclidean* antara data dan *medoids* baru
Apakah Total jarak (S) > 0?
Tidak
Ya
Hasil pembentukan *cluster*
a



Gambar 1.1 *Flowchart* Penelitian