PENERAPAN ALGORITMA *K-MEDOIDS* DENGAN OPTIMASI *GAP STATISTICS* DALAM PENGELOMPOKAN DAERAH RAWAN KRIMINALITAS DI INDONESIA

OKTAVIA DYAHERAWATI
H1091211011

SKRIPSI

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

PENERAPAN ALGORITMA *K-MEDOIDS* DENGAN OPTIMASI *GAP STATISTICS* DALAM PENGELOMPOKAN DAERAH RAWAN KRIMINALITAS DI INDONESIA

OKTAVIA DYAHERAWATI
H1091211011

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 Algoritma *K-Medoids* dengan Optimasi *Gap Statistics* dalam Pengelompokan Daerah Rawan Kriminalitas di Indonesia**
Nama Mahasiswa	: Oktavia Dyaherawati
NIM	: H1091211011
Jurusan/Program Studi	: Matematika/Statistika
Tanggal Lulus	: 14 Maret 2025
SK Pembimbing	: No.2590/UN22.8/TD.06/2024 /Tanggal 2 September 2024
SK Penguji	: No.803/UN22.8/TD.06/2025 /Tanggal 10 Maret 2025

Dosen Pembimbing

Pembimbing I	Pembimbing II
Shantika Martha, M.Si	Nurfitri Imro'ah, M.Si
NIP 198403082008122003	NIP 198907182019032021

Dosen Penguji

Ketua Penguji	Anggota Penguji
Dr. Evy Sulistianingsih, M.Sc	Hendra Perdana, M.Sc
NIP 198502172008122006	198810102019031020
Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Shantika Martha, M.Si	Nurfitri Imro'ah, M.Si
NIP 198403082008122003	NIP 198907182019032021

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 : Oktavia Dyaherawati

NIM : H1091211011

Program Studi/ Jurusan : Statistika/Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyatan ini untuk digunakan sebagaimana mestinya.

Pontianak, 14 Maret 2025

Oktavia Dyaherawati
NIM H1091211011

Penerapan Algoritma *K-Medoids* dengan Optimasi *Gap Statistics* dalam Pengelompokan Daerah Rawan Kriminalitas di Indonesia

Abstrak

Banyaknya keinginan yang tidak diiringi dengan kemampuan dan kesanggupan dalam memenuhi kebutuhan hidup menjadikan seseorang melakukan berbagai cara untuk mendapatkan kebutuhan materi, seperti melakukan tindakan kejahatan. Informasi terkait banyaknya tindak kejahatan yang terjadi sangat dibutuhkan oleh masyarakat dan penegak hukum. Penelitian ini bertujuan untuk mengelompokkan daerah rawan kriminalitas pada provinsi di Indonesia menggunakan algoritma *K-Medoids* dengan optimasi yang digunakan yaitu *Gap Statistics*. Algoritma *K-Medoids* merupakan metode untuk mengatasi permasalahan pada *K-Means clustering* yang sensitif terhadap *outlier* dengan menggunakan perwakilan dari objek sebagai pusat *cluster*. Penentuan jumlah *cluster* terbaik pada metode ini masih belum memiliki dasar teori yang jelas, sehingga diperlukan pendekatan untuk mengidentifikasi jumlah *cluster* optimal. *Gap statistics* merupakan salah satu pendekatan terbaik untuk menentukan jumlah *cluster* optimal dengan membangkitkan data acak dalam penentuan jumlah kelompok optimum. Data yang digunakan merupakan data sekunder yang didapatkan dari publikasi Badan Pusat Statistik yaitu Statistik Kriminal 2023 yang berisi data jumlah kriminalitas menurut jenis kejahatan dan kepolisian daerah tahun 2022. Penelitian ini berfokus untuk membentuk kelompok yang berisi provinsi dengan jarak terdekat berdasarkan karakteristik dari kriminalitas menggunakan analisis *cluster*. Berdasarkan analisis yang telah dilakukan, jumlah *cluster* optimal yang terbentuk berjumlah empat *cluster* dengan nilai *gap statistics* yang diperoleh sebesar 0,65. *Cluster* 1 dikategorikan sebagai daerah sangat rawan kriminalitas yang terdiri dari tiga provinsi, *Cluster* 2 dikategorikan sebagai daerah rawan kriminalitas yang terdiri dari dua provinsi, *Cluster* 3 dikategorikan sebagai daerah cukup rawan kriminalitas yang terdiri dari tujuh provinsi, dan *Cluster* 4 dikategorikan sebagai daerah tidak rawan kriminalitas yang terdiri dari 22 provinsi.

Kata Kunci: Analisis *Cluster*, *Cluster* Optimal, *Outlier*

Application of K-Medoids Algorithm with Gap Statistics Optimization in Clustering Crime Prone Areas in Indonesia

Abstract

The number of desires that are not accompanied by the ability and ability to meet the needs of life makes a person do various ways to get material needs, such as committing crimes. Information related to the number of crimes that occur is needed by the community and law enforcement. This study aims to classify crime-prone areas in provinces in Indonesia using the K-Medoids algorithm with the optimization used, namely Gap Statistics. The K-Medoids algorithm is a method to overcome problems in K-Means clustering that are sensitive to outliers by using representatives of objects as cluster centers. Determining the best number of clusters in this method still does not have a clear theoretical basis, so an approach is needed to identify the optimal number of clusters. Gap statistics is one of the best approaches to determine the optimal number of clusters by generating random data in determining the optimum number of groups. The data used is secondary data obtained from the Central Bureau of Statistics publication, namely Criminal Statistics 2023 which contains data on the number of crimes by type of crime and regional police in 2022. This research focuses on forming groups containing provinces with the closest distance based on the characteristics of crime using cluster analysis. Based on the analysis that has been done, the optimal number of clusters formed is four clusters with a gap statistics value obtained of 0.65. Cluster 1 is categorized as a very crime-prone area consisting of three provinces, Cluster 2 is categorized as a crime-prone area consisting of two provinces, Cluster 3 is categorized as a moderately crime-prone area consisting of seven provinces, and Cluster 4 is categorized as a non-crime-prone area consisting of 22 provinces.

Keywords: *Cluster Analysis, Optimal Cluster, Outliers*

PRAKATA

Assalamu'alaikum Warahmatullahi Wabarakatuh

Alahmdulillahirabbil'alamin, puji syukur kehadirat Allah SWT karena atas segala berkah, rahmat serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Penerapan Algoritma *K-Medoids* dengan Optimasi *Gap Statistics* dalam Pengelompokan Daerah Rawan Kriminalitas di Indonesia"** sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika (S.Stat) pada Program Studi Statistika di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Tanjungpura.

Dalam penulisan skripsi ini, begitu banyak hal yang telah penulis hadapi untuk dapat menyelesaikan skripsi ini, terdapat banyak dukungan, bimbingan, serta motivasi dan doa dari berbagai pihak sehingga penulis dapat menyelesaikan penulisan skripsi ini dengan lancar. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua, abang-abang, beserta keluarga tercinta yang selalu mendoakan dan memberikan semangat. motivasi, serta dukungan moril maupun materil kepada penulis selama masa perkuliahan maupun dalam menyelesaikan skripsi ini.
2. Ibu Dr. Evi Noviani, M.Si. selaku dosen pembimbing akademik yang telah memberikan saran, masukan dan motivasi kepada penulis selama masa perkuliahan.
3. Ibu Shantika Martha, M.Si. selaku dosen pembimbing pertama yang telah meluangkan waktu untuk membimbing, mengarahkan serta memberikan masukan dan motivasi kepada penulis dalam menyelesaikan skripsi ini.
4. Ibu Nurfitri Imro'ah, M.Si. selaku dosen pembimbing kedua yang telah meluangkan waktu untuk membimbing, mengarahkan serta memberikan masukan dan motivasi kepada penulis dalam menyelesaikan skripsi ini.
5. Ibu Dr. Evy Sulistianingsih, M.Sc. selaku dosen penguji pertama yang telah memberikan saran, masukan, serta koreksi dalam penyempurnaan penulisan skripsi ini.

6. Bapak Hendra Perdana, M.Sc. selaku dosen penguji kedua yang telah memberikan saran, masukan, serta koreksi dalam penyempurnaan penulisan skripsi ini.
7. Teman-teman terdekat penulis, Muthiara, Putri, Della, Wahida, Lusi, Ulya, Siska, Nanda, Ghina dan Jihan yang telah membersamai langkah penulis di masa perkuliahan dan selalu memberikan semangat, motivasi, serta selalu bersama-sama melewati masa perkuliahan baik senang maupun susah
8. Dhimas, Mira, Diah, Cindy, dan Ola yang telah seperti saudara bagi penulis di kota rantauan dengan selalu menemani, membantu, dan menjadi tempat cerita serta berbagai canda tawa yang membahagiakan bagi penulis.
9. Teman-teman Expos 21 yang telah membersamai langkah dan cerita penulis di masa perkuliahan serta memberikan banyak semangat kepada penulis.
10. Semua pihak yang tidak dapat disebutkan satu-persatu yang telah memberikan dukungan dan bantuan kepada penulis.
11. Dan yang terakhir, kepada diri saya sendiri, Oktavia Dyaherawati. Apresiasi yang sebesar-besarnya karena telah bertanggung jawab menyelesaikan apa yang telah di mulai. Terima kasih sudah terus berjuang dan tidak mengenal kata menyerah. Terima kasih sudah membawa nama ini terpatri indah dalam skripsi. Berbahagialah dimanapun berada, Hera. Apapun kurang dan lebihmu mari merayakan diri sendiri

Semoga seluruh kebaikan, dukungan, bantuan dan motivasi yang telah diberikan mendapatkan balasan pahala yang setimpal dari Allah SWT. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, maka dari itu penulis mengharapkan saran dan kritik yang membangun untuk kesempurnaan skripsi ini. Semoga skripsi ini dapat bermanfaat bagi tiap pembacanya.

Pontianak, 14 Maret 2025

Oktavia Dyaherawati

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 5
BAB II LANDASAN TEORI 11
2.1 Kriminalitas 11
2.2 Analisis Multivariat 12
2.3 Analisis *Cluster* 13
2.3.1 Metode *Hierarchical Clustering* 13
2.3.2 Metode *Non-Hierarchical Clustering* 14
2.4 Asumsi Analisis *Cluster* 15
2.5 Standarisasi *Z-Score* 17
2.6 *Euclidean Distance* 17
BAB III ALGORITMA *K-MEDOIDS* DAN *GAP STATISTICS* 19
3.1 Algoritma *K-Medoids* 19
3.2 *Gap Statistics* 20
BAB IV HASIL DAN PEMBAHASAN 23
4.1 Deskripsi Data 23

4.2 Asumsi Analisis Cluster 25
4.3 Standarisasi Data 26
4.4 Penentuan Jumlah *Cluster* Awal 27
4.5 *Clustering* Menggunakan Algoritma *K-Medoids* 27
4.6 Validasi Hasil *Cluster* Menggunakan *Gap Statistics* 31
4.7 Profilisasi Hasil *Cluster* 34
BAB V PENUTUP 39
5.1 Kesimpulan 39
5.2 Saran 40
DAFTAR PUSTAKA 41
LAMPIRAN 44

DAFTAR TABEL

Tabel 4.1 Variabel Penelitian 23
Tabel 4.2 Statistik Deskriptif.....24
Tabel 4.3 Nilai *Variance Inflation Factor* (VIF) 25
Tabel 4.4 Hasil *Variance Inflation Factor* (VIF) Tanpa X_6.....26
Tabel 4.5 Hasil *Variance Inflation Factor* (VIF) Tanpa X_9.....26
Tabel 4.8 Pengelompokan Provinsi di Indonesia 30
Tabel 4.6 Nilai rata-rata jumlah kuadrat jarak pada empat *cluster* 32
Tabel 4.7 Nilai *Gap Statistics*.....33
Tabel 4.9 Karakteristik Pengelompokan algoritma *K-Medoids* 35

DAFTAR GAMBAR

Gambar 1.1 *Flowchart* Penelitian.....10
Gambar 4.1 Grafik *Cluster* Optimal dengan Metode *Gap Statistics*.....33
Gambar 4.2 Peta Hasil Pengelompokan Daerah Rawan Kriminalitas.....37

DAFTAR LAMPIRAN

Lampiran I Data Jumlah Kejahatan Menurut Jenis Kejahatan dan Kepolisian Daerah Tahun 2022 43
Lampiran II Data Hasil Standarisasi 44
Lampiran III Jarak *Euclidean Medoids* Awal 45
Lampiran IV Jarak *Euclidean Medoids* Baru 46

DAFTAR SIMBOL

p : banyaknya variabel

n : banyaknya objek

r : banyaknya *cluster*

r_{jl} : koefisien korelasi antar variabel ke-*j* dan variabel ke-*l*, dengan *l* = 1, 2,...,*p*

a_{jlo} : koefisien korelasi parsial antar variabel ke-*j* dan variabel ke-*l*, dengan variabel *o* konstan

$\bar{x}_j$: rata-rata variabel ke-*j*, dengan *j* = 1, 2,…,*p*

$\bar{x}_l$: rata-rata variabel ke-*l*, dengan *l* = 1, 2,…,*p*

R_j^2 : koefisien determinasi antara X_j dengan variabel bebas lainnya pada model dugaan, *j*: 1, 2, ..., *p*

Z_{ij} : standarisasi objek ke-*i* variabel ke-*j*, dengan *i* = 1, 2, …, *n* dan *j* = 1, 2,…,*p*

x_{ij} : data ke-*i* pada variabel ke-*j*

σ_j : standar deviasi variabel ke-*j*

d_{ik} : jarak *eulidean* antara objek ke-*i* ke pusat *cluster* ke-*k*. dengan *k* = 1,2,…,*r*

c_{kj} : nilai pusat *cluster* ke-*k* pada variabel ke-*j*, dengan *k* = 1,2, …, *r*

Dr : jumlah jarak berpasangan antar titik dalam *cluster*

W_k : rata-rata jumlah kuadrat jarak dalam *cluster* ke-*k*

n_k : banyaknya anggota *cluster* ke-*k*

$Gap(K)$: nilai *gap statistics* pada *K cluster*

B : *resampling* dari data acak

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan ilmu pengetahuan dan teknologi di era globalisasi menimbulkan berbagai persaingan di kehidupan manusia dalam hal etika, moral, ideologi, sosial, dan ekonomi (Putra, Martha, Fikram, dan Yuhan., 2020). Hal ini berdampak pada perubahan yang ada di dalam masyarakat. Perubahan yang terjadi disebabkan oleh nilai-nilai yang terdapat dalam masyarakat, sehingga berdampak pada perilaku manusia. Keluar masuknya budaya-budaya asing di Indonesia dengan membawa nilai-nilai baru menjadi salah satu penyebab terjadinya perubahan di era globalisasi (Rahmalia, Ariusni, Triani., 2019). Munculnya sifat materialisme, hedonisme dan lain sebagainya menjadikan manusia selalu memiliki keinginan yang berlebih dalam pemenuhan kebutuhan hidup. Banyaknya keinginan yang tidak diiringi dengan kemampuan dan kesanggupan dalam memenuhi kebutuhan hidup menjadikan seseorang melakukan berbagai cara untuk mendapatkan kebutuhan materi, seperti melakukan tindakan kejahatan (Dermawanti, Hoyyi, dan Rusgiyono., 2015).

Tindak kejahatan atau kriminalitas bukanlah suatu hal baru di Indonesia. Telah banyak kasus kejahatan yang menimbulkan kerugian fisik dan harta orang lain. Adapun jenis kriminalitas yang membahayakan fisik yaitu berupa kekerasan, pelecehan, pembunuhan, narkotika dan obat-obatan terlarang. Sedangkan kriminalitas yang dapat mengancam harta orang lain yaitu berupa korupsi, penipuan, penggelapan dana, dan perampasan hak milik (Riyardi & Guritno, 2022). Meningkatnya jumlah kejadian kejahatan di Indonesia menjadi suatu ancaman besar bagi rasa aman untuk berkehidupan ditengah perubahan yang terjadi pada nilai-nilai kemasyarakatan. Tercipta dan terpenuhinya keamanan akan membangun suasana yang kondusif bagi masyarakat dalam melakukan berbagai aktivitas. Rasa aman menjadi hak asasi manusia dalam berkehidupan, sehingga baik pemerintah maupun masyarakat harus saling bekerja sama untuk mengurangi kejahatan di Indonesia (Inayah, Maghfiroh, dan Novitasari., 2022).

Maraknya tindak kejahatan menjadi salah satu tantangan terbesar bagi pemerintah Indonesia dalam menjaga keamanan dan meningkatkan kesejahteraan masyarakat. Berdasarkan data dari registrasi Polri yang dipublikasikan oleh Badan Pusat Statistik (BPS), jumlah kejahatan di Indonesia menunjukkan fluktuasi dalam tiga tahun terakhir. Pada tahun 2020, total kejadian kejahatan yang terjadi mencapai 247.218 kejadian, kemudian pada tahun 2021 mengalami penurunan menjadi 239.481 kejadian. Namun, penurunan ini tidak berlangsung lama karena pada tahun 2022 terjadi peningkatan yang mencapai 372.965 kejadian. Hal serupa juga terjadi pada tingkat kejahatan (*crime rate*), yang mengalami peningkatan pada tahun 2022 sebesar 137 per 100.000 penduduk. Ini berarti dari 100.000 penduduk, terdapat 137 orang yang menjadi korban kejahatan. Selain itu, pada tahun 2022 juga terjadi penurunan dalam interval waktu terjadinya kejahatan (*crime clock*), yang menjadi 00.01’24” (1 menit 24 detik). Penurunan interval waktu terjadinya kejahatan ini menunjukkan peningkatan frekuensi kejadian tindak kejahatan (Badan Pusat Statistik, 2023).

Meningkatnya jumlah kejadian kejahatan di Indonesia yang didasarkan pada jenis kejahatan seperti kejahatan terhadap nyawa, kejahatan terhadap fisik/badan, kejahatan terhadap kesusilaan, dan masih banyak lagi. Tindak kejahatan dapat terjadi dimanapun dan kapanpun, sehingga sulit dalam menentukan daerah mana yang memiliki tingkat kerawanan kriminalitas. Informasi terkait banyaknya tindak kejahatan yang terjadi sangat dibutuhkan oleh masyarakat dan penegak hukum seperti jajaran kepolisian. Dengan adanya informasi tersebut, masyarakat dapat mengantisipasi hal-hal yang akan terjadi, khususnya bagi kepolisian dapat membantu dalam pengambilan keputusan apakah suatu daerah memerlukan pengawasan ekstra atau tidak (Insiyah, Khasanah, dan Hendarsyah., 2023).

Banyaknya kejadian kejahatan di Indonesia menjadikannya perlu dilakukan pengelompokan daerah-daerah rawan kriminalitas menurut jenis kejahatan dan kepolisian daerah tahun 2022, sehingga pemerintah dapat memperketat sistem keamanan dan masyarakat dapat lebih waspada terhadap kejahatan yang terjadi di daerah tersebut. Dikarenakan banyak kejadian kejahatan yang terjadi, maka untuk menganalisis kasus ini diperlukan analisis *cluster*. Analisis *cluster* bertujuan untuk

mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Dalam hal ini, analisis *cluster* mengklasifikasi suatu objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain akan berada dalam *cluster* yang sama (Ediyanto, Mara, dan Satyahadewi., 2013).

Penelitian ini berfokus untuk membentuk kelompok yang berisi provinsi dengan jarak terdekat berdasarkan karakteristik dari kriminalitas menggunakan analisis *cluster* dengan algoritma *K-Medoids. K-Medoids clustering* atau dikenal sebagai *Partioning Around Medoids (PAM)* adalah varian dari metode *K-Means. K-Medoids* digunakan untuk mengatasi kelemahan pada *K-Means clustering* yang sensitif terhadap *outlier*. Dalam metode *K-Medoids* akan dipilih satu objek representatif yang digunakan untuk mewakili *cluster* (Okta, Satyahadewi, dan Debataraja., 2019). Penentuan jumlah *cluster* membutuhkan metode optimasi dimana penentuan *cluster* optimal merupakan hal yang perlu diperhatikan dalam analisis *cluster* untuk memperoleh partisi yang paling sesuai dengan data. Pada penelitian ini, optimasi yang digunakan yaitu *gap statistics*. *Gap statistics* merupakan optimasi dengan membangkitkan data acak dari distribusi *referensi* dalam penentuan jumlah kelompok optimum. Kelebihan dari validasi *gap statistics* yaitu membentuk *cluster* dengan hasil yang lebih baik dengan mengatasi keterbatasan pada metode *elbow*, *silhouette*, dan *hartigan statistics*, dikarenakan pada optimasi ini akan membandingkan hasil dari data asli terhadap data acak (Yang, Lee, Choi, dan Joo., 2019). Oleh karena itu, pada penelitian ini digunakan metode *K-Medoids* dengan optimasi *gap statistics* untuk membentuk kelompok terbaik. Dari kelompok yang terbentuk, dapat dilihat daerah-daerah yang termasuk ke dalam daerah rawan kriminalitas, sehingga pemerintah dan aparat kepolisian dapat memprioritaskan daerah-daerah yang termasuk ke dalam kelompok rawan kriminalitas.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, diperoleh rumusan masalah yaitu bagaimana pengelompokan daerah rawan kriminalitas pada provinsi di Indonesia menggunakan algoritma *K-Medoids* dengan optimasi *Gap Statistics*?

1.3 Tujuan Penelitian

Berdasarkan permasalahan yang telah dipaparkan, maka tujuan dari penelitian ini yaitu untuk mengelompokkan daerah rawan kriminalitas pada provinsi di Indonesia menggunakan algoritma *K-Medoids* dengan optimasi *Gap Statistics*.

1.4 Batasan Masalah

Batasan masalah dalam penelitian ini yaitu:

1. Data yang digunakan merupakan data sekunder yang didapatkan dari publikasi Badan Pusat Statistik (BPS) yaitu Statistik Kriminal 2023. Data tersebut terkumpul per tahun 2022 pada setiap Provinsi di Indonesia
2. Terdapat sembilan variabel yang digunakan yaitu kejahatan terhadap nyawa; kejahatan terhadap fisik/badan; kejahatan terhadap kesusilaan; kejahatan terhadap kemerdekaan orang; kejahatan terhadap hak milik/barang dengan penggunaan kekerasan; kejahatan terhadap hak milik/barang tanpa penggunaan kekerasan; kejahatan terkait narkotika, kejahatan terkait penipuan, penggelapan, dan korupsi; dan kejahatan terhadap ketertiban umum.
3. *Cluster* optimal yang dibentuk maksimal berjumlah 10.
4. Perhitungan jarak menggunakan rumus jarak *Euclidean.*

1.5 Tinjauan Pustaka

Nahdliyah, Widiharih, dan Prahutama (2019) telah melakukan penelitian yang bertujuan untuk membandingkan hasil dari pengukuran jarak yaitu *euclidean* dan *cityblock* pada metode *K-Medoids* dengan validasi *Silhouette Index* dan *C-Index* pada kasus jumlah kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018. Hasil dari penelitian menunjukkan bahwa *cluster* yang optimal pada k = 4 dengan jarak *euclidean* dimana nilai SI = 0,3862593 dan CI = 0,043893. Berdasarkan hasil pengklasteran pada metode ini didapatkan bahwa jarak pengukuran yang digunakan akan berpengaruh terhadap hasil pengklasteran.

K-Medoids merupakan salah satu metode yang digunakan untuk proses *clustering.* Penelitian bertujuan untuk menentukan banyaknya jumlah *cluster* optimal pada analisis *K-Medoids* dengan menggunakan metode *Sillhoutte Coefficient*, metode *Elbow*, dan metode *gap statistics* berdasarkan nilai rata-rata

dari *Dunn Index*. Hasil penelitian menunjukkan bahwa metode *gap statistics* memiliki nilai rata-rata *Dunn Index* paling besar dari data yang dibangkitkan untuk data negatif yang di nol kan yaitu senilai 0,125734 dan metode *Sillhouette Coefficient* memiliki nilai rata-rata *Dunn Index* paling besar untuk data negatif yang dimutlakkan yaitu senilai 0,113315. Berdasarkan hasil penelitian tersebut, dapat disimpulkan bahwa penentuan jumlah *cluster* optimal dengan metode yang berbeda dipengaruhi oleh objek dan jumlah data yang digunakan (Ramadhania, Widiarti, Zakaria, dan Nusyirawan., 2023)

Hoerunnisa, Dwilestari, Dikananda, Sunana, dan Pratama (2024) melakukan pengelompokan dengan komparasi algoritma *K*-Means dan K-*Medoids* dalam analisis pengelompokan daerah rawan kriminalitas di Indonesia. Pengelompokan ini melibatkan penentuan *Davies Bouldin Index* (DBI) untuk tiap *cluster*. Hasil yang diperoleh pada penelitian ini yaitu pada proses *K-Means* menghasilkan 5 *cluster*, sementara *K-Medoids* menghasilkan 2 *cluster*. Hasil perbandingan menggunakan DBI menunjukkan bahwa *K-Means* lebih optimal (DBI 0,463) dibandingkan K-Medoids (DBI 1,089).

Tampubolon, Suhada, Safii, Solikhun, dan Suhendro (2021) melakukan pengelompokan menggunakan Algotitma *K*-Means dan K-*Medoids Clustering* pada kasus tindak kriminalitas di Indonesia berdasarkan provinsi dari tahun 2017-2019 dengan jumlah data sebanyak 34 provinsi. Hasil yang diperoleh yaitu pada algoritma *K-Means* diperoleh hasil *cluster* 1 dengan kategori tindak kriminalitas tinggi memiliki 6 anggota, *cluster* 2 dengan kategori tindak kriminalitas rendah memiliki 28 anggota. Sedangkan hasil perhitungan manual algoritma *K-Medoids* diperoleh hasil *cluster* 1 dengan kategori tindak kriminalitas tinggi memiliki 7 anggota, cluster 2 dengan kategori tindak kriminalitas rendah memiliki 27 anggota.(Tampubolon et al., 2021)

1.6 Metodologi Penelitian

Penelitian ini menggunakan data sekunder yang didapatkan dari publikasi Badan Pusat Statistik. Banyaknya observasi yang digunakan yaitu sesuai dengan banyaknya provinsi yang ada di Indonesia yaitu sebanyak 34 provinsi dengan 9 variabel yang merupakan jenis kejahatan yang terjadi. Data yang digunakan diolah

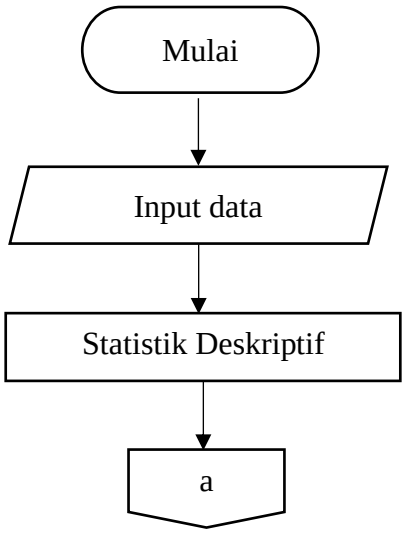
menggunakan *software Microsoft Excel* dan RStudio. Penelitian ini dimulai dengan studi literatur terkait kriminalitas di Indonesia dan dilanjutkankan dengan penentuan variabel yang berkaitan dengan kriminalitas. Kemudian dilanjutkan dengan mempelajari dan memahami terkait metode yang digunakan yaitu algoritma *K-Medoids.*

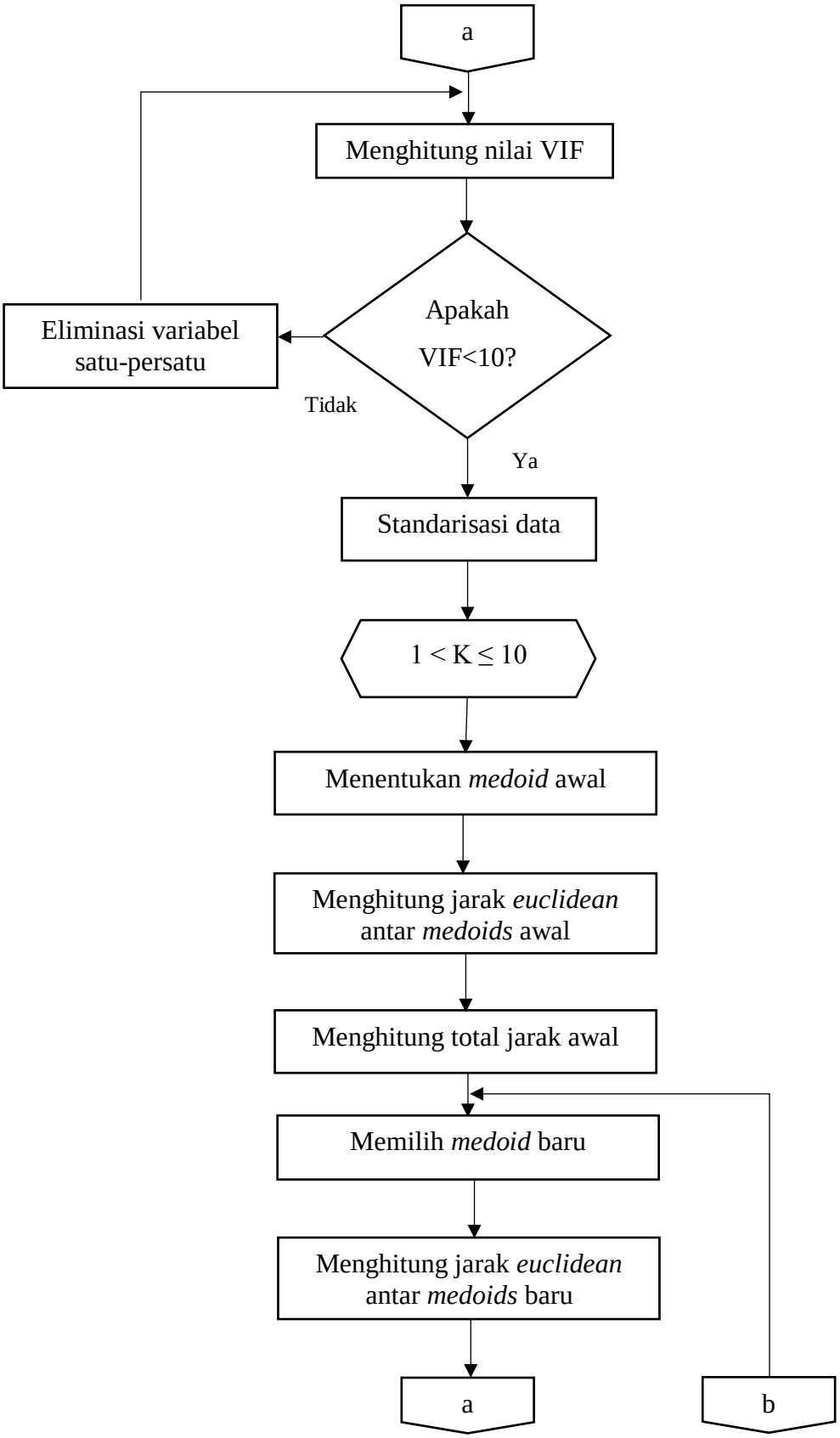
Pada tahapan pertama, dilakukan penginputan data yang digunakan dalam penelitian. Terdapat dua jenis data yang diiput yaitu data objek (provinsi) dan data subjek (variabel). Pada penelitian ini, data objek yang digunakan akan membentuk kelompok dari provinsi di Indonesia berdasarkan kesamaan dari variabel yang digunakan. Setelah dilakukan penginputan data, dilakukan analisis statistik deskriptif untuk melihat ringkasan pada setiap variabel yang digunakan. Kemudian dilakukan pengujian multikolinearitas untuk melihat hubungan linear diantara semua variabel dengan menggunakan nilai *Variance Inflation Factor* (VIF). Multikolinearitas dapat terjadi apabila nilai VIF yang didapatkan pada tiap variabel lebih dari sama dengan 10, sehingga perlu dilakukan eliminasi untuk variabel dengan nilai VIF lebih dari sama dengan 10 tersebut. Namun, jika nilai VIF pada semua variabel kurang dari 10, maka dapat dilanjutkan ke proses berikutnya yaitu standarisasi data. Standarisasi data dilakukan dengan menghitung nilai *z-score* dengan tujuan untuk menyamakan satuan antar tiap variabel serta meningkatkan keakuratan dan konsistensi pada hasil. Setelah dilakukan proses standarisasi data, maka tahap selanjutnya yaitu menentukan *k* sebagai banyaknya klaster yang ingin dibentuk dan membangkitkan *k* pusat klaster (*medoid*) secara acak. Selanjutnya yaitu menghitung jarak antar objek *non-medoid* dengan *medoid* pada tiap *cluster* dan menempatkan tiap objek ke jarak terdekat, kemudian akan dihitung total jaraknya. Dalam hal ini, jarak yang digunakan yaitu *euclidean distance.* Tahap selanjutnya yaitu tahapan iterasi dengan memilih secara acak objek pada masing-masing *cluster* sebagai *medoid* baru. Kemudian dihitung jarak setiap objek yang terdapat pada masing-masing *cluster* dengan kandidat *medoid* baru dan hitung total jaraknya. Tahap selanjutnya yaitu menghitung total simpangan (*S*) dengan menghitung nilai total jarak baru – nilai total jarak lama. Jika nilai $S < 0$, maka objek akan ditukar dengan data *cluster* sehingga membentuk sekumpulan *k* objek

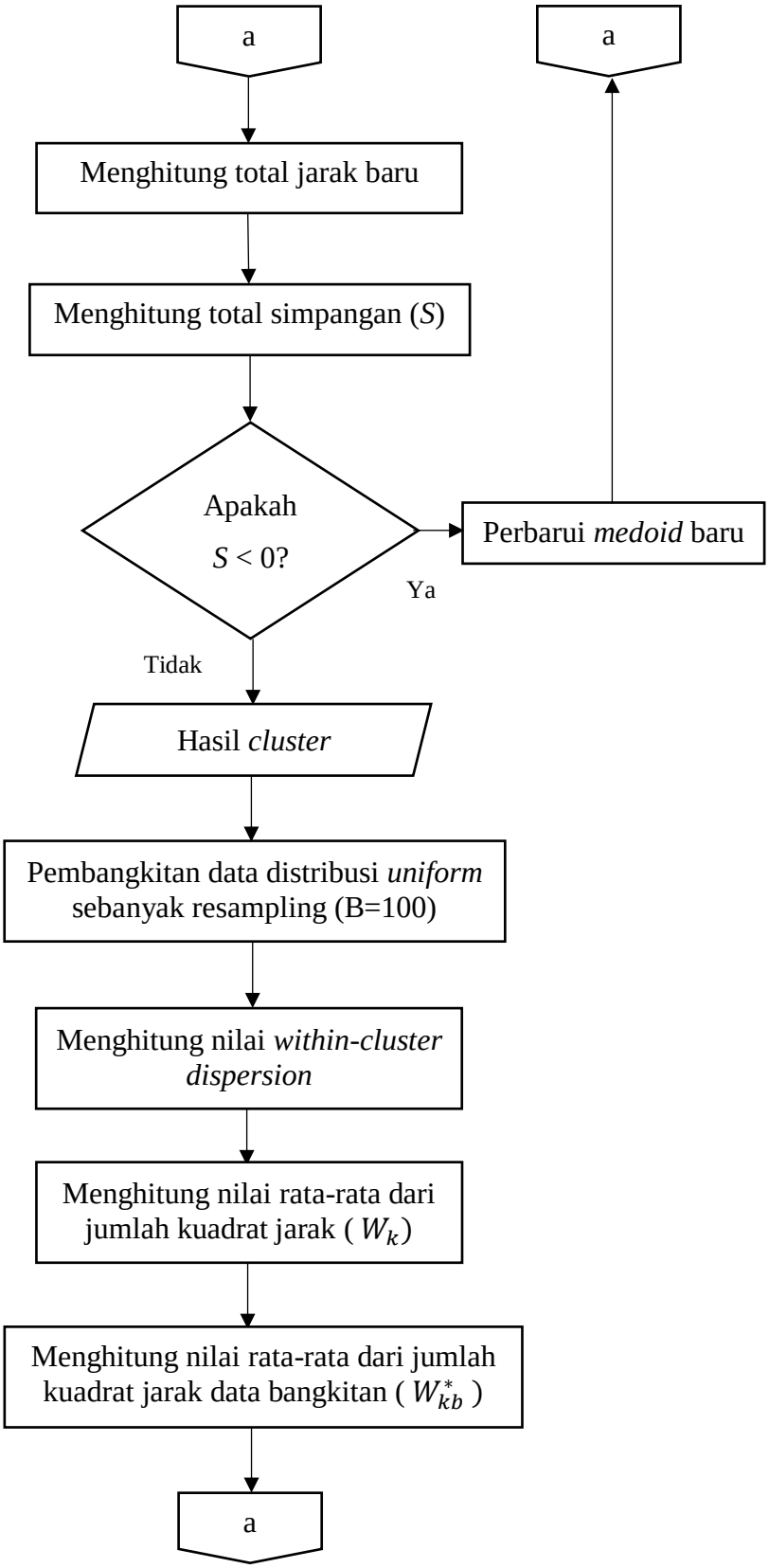
sebagai *medoid* baru. Jika hal tersebut terjadi, maka tahap selanjutnya yaitu mengulangi ke langkah memilih secara acak objek pada masing-masing *cluster* sebagai *medoid* baru hingga tidak terjadi perubahan lagi. Proses iterasi akan berhenti apabila nilai S > 0 dan pada tahap ini akan diperoleh hasil klaster beserta anggota masing-masing *cluster.*

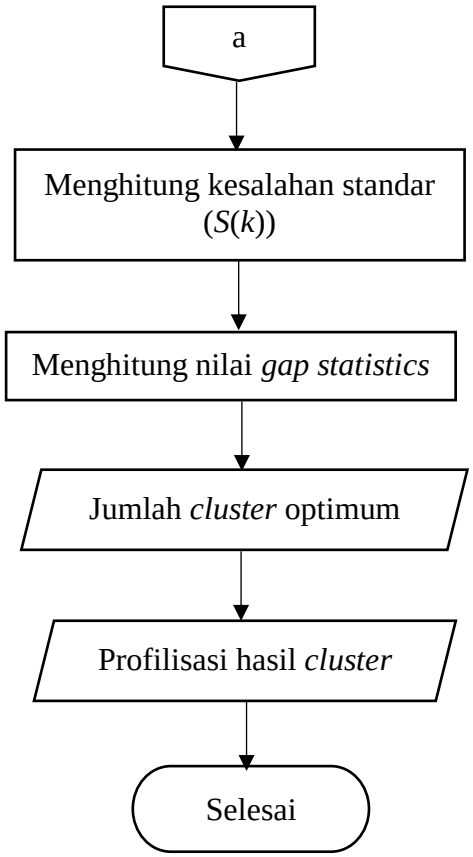
Pada tahap akhir penelitian dilakukan validasi terhadap hasil analisis *cluster* untuk memperoleh partisi yang paling sesuai dengan data. Jika suatu *cluster* tidak dilakukan proses validasi, maka akan berpengaruh pada hasil analisis. Pada penelitian ini digunakan validasi *gap statistics*, dimana pada validasi perlu membangkitkan data distribusi referensi dalam penentuan jumlah kelompok optimum. Metode *gap statistics* akan membentuk *cluster* dengan hasil yang lebih konstan yaitu dengan membandingkan hasil dari data asli terhadap data acak. Selanjutnya, dihitung nilai *gap statistics* berdasarkan jumlah *cluster* yang telah ditentukan dengan menghitung selisih antara rata-rata nilai log dari data acak dan log dari data asli. Jumlah *cluster* optimal dapat dilihat dari nilai *gap* tertinggi berdasarkan grafik. Setelah diperoleh jumlah *cluster* optimal, maka dilakukan profilisasi hasil *cluster* dan membuat kesimpulan dari hasil yang diperoleh.

Tahapan yang dilakukan dalam penelitian ini dengan menggunakan algoritma *K-Medoids* digambarkan melalui diagram alir yang dapat dilihat pada Gambar 1.1 berikut




a
Menghitung nilai VIF
Apakah
VIF<10?
Eliminasi variabel
satu-persatu
Tidak
Ya
Standarisasi data
1 < K ≤ 10
Menentukan *medoid* awal
Menghitung jarak *euclidean*
antar *medoids* awal
Menghitung total jarak awal
Memilih *medoid* baru
Menghitung jarak *euclidean*
antar *medoids* baru
a
b


a
a
Menghitung total jarak baru
Menghitung total simpangan (S)
Apakah
S < 0?
Perbarui medoid baru
Ya
Tidak
Hasil cluster
Pembangkitan data distribusi uniform sebanyak resampling (B=100)
Menghitung nilai within-cluster dispersion
Menghitung nilai rata-rata dari jumlah kuadrat jarak (W_k)
Menghitung nilai rata-rata dari jumlah kuadrat jarak data bangkitan (W_{kb}^*)
a



Gambar 1.1 Flowchart Penelitian