

**ANALISIS *CLUSTER ROBUST* MENGGUNAKAN METODE *K-MEDOIDS*
PADA DATA MENGANDUNG PENCILAN**

(Skripsi)

Oleh

**EPMI ANNISA NUR FAJRINA
NPM 1757031006**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021**

ABSTRACT

ANALYSIS CLUSTER ROBUST USING K-MEDOIDS METHOD IN THE DATA CONTAINING OUTLIER

By

EPMI ANNISA NUR FAJRINA

Cluster analysis is a multivariate analysis technique that is useful for grouping observational data or variables into clusters in such a way that each cluster is homogeneous according to the factors used for clustering. There are two assumptions that must be met in cluster analysis, namely a representative sample and non-multicollinearity between each variable. In fact, not all data meet the two assumptions above, especially for the assumption of a representative sample because of the problem of outliers. To overcome the problem of outliers, there is one method of non-hierarchical cluster analysis that is resistant to outliers, namely K-Medoids. K-Medoids method including the sealing technique (partition) which divides or separates objects to different groups so that objects with the same characteristics are in the same cluster and objects with different characteristics are in another clusters. This study aims to examine robust cluster analysis with the K-medoids method through simulation data with several percentages of outlier values and see the effectiveness of the K-medoids method to overcome outliers in terms of the sillhouette index and Dunn index validation values, and it is concluded that the analysis the cluster K-medoids method is an effective clustering method used on data that contains outliers. Based on the average value of the sillhouette index and Dunn index, the results show that the K-medoids method is effectively used for clustering data with a small number of objects, a small number of clusters, and a large proportion of outliers.

Keywords: Outliers, Cluster Analysis, K-Medoids, sillhouette index, Dunn index

ABSTRAK

ANALISIS *CLUSTER ROBUST* MENGGUNAKAN METODE *K-MEDOIDS* PADA DATA MENGANDUNG PENCILAN

Oleh

EPMI ANNISA NUR FAJRINA

Analisis kluster merupakan suatu teknik analisis multivariat yang berguna untuk mengelompokkan data observasi ataupun variabel-variabel ke dalam kluster sedemikian rupa sehingga masing-masing kluster bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran. Ada dua asumsi yang harus dipenuhi pada analisis kluster, yaitu sampel representatif dan tidak ada multikolinearitas antara tiap variabel. Pada kenyataannya, tidak semua data memenuhi kedua asumsi di atas, khususnya untuk asumsi sampel representatif karena adanya masalah pencilan. Untuk mengatasi masalah pencilan tersebut, terdapat salah satu metode analisis kluster non hierarki yang tahan terhadap pencilan yaitu *K-Medoids*. Metode *K-Medoids* termasuk dalam teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke-*k* kelompok yang berbeda sehingga objek yang memiliki karakteristik yang sama masuk ke dalam satu kluster yang sama dan objek yang mempunyai karakteristik berbeda masuk ke dalam kluster yang lain. Penelitian ini bertujuan untuk mengkaji analisis kluster *robust* dengan metode *K-medoids* melalui data simulasi dengan beberapa persentase nilai pencilan dan melihat keefektifan metode *K-medoids* untuk mengatasi pencilan ditinjau dari nilai validasi *sillhouette index* dan *Dunn index*, dan diperoleh kesimpulan bahwa analisis kluster metode *K-medoids* merupakan metode pengklasteran yang efektif digunakan pada data yang mengandung pencilan. Berdasarkan rata-rata nilai *sillhouette index* dan *Dunn index* diperoleh hasil bahwa metode *K-medoids* efektif digunakan untuk pengklasteran pada data dengan jumlah objek kecil, jumlah kluster kecil, dan proporsi pencilan yang besar.

Kata kunci: Pencilan, Analisis Kluster, *K-Medoids*, *sillhouette index*, *Dunn index*

**ANALISIS *CLUSTER ROBUST* MENGGUNAKAN METODE *K-MEDOIDS*
PADA DATA MENGANDUNG PENCILAN**

Oleh

EPMI ANNISA NUR FAJRINA

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021**

Judul Skripsi : **ANALISIS CLUSTER ROBUST MENGGUNAKAN METODE K-MEDOIDS PADA DATA MENGANDUNG PENCILAN**

Nama Mahasiswa : **Epmi Annisa Nur Fajrina**

Nomor Pokok Mahasiswa : **1757031006**

Program Studi : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**



1. **Komisi Pembimbing**

Dr. Khoirin Nisa, M.Si.
NIP. 197407262000032001

Subian Saidi, S.Si., M.Si.
NIP. 198008212008121001

2. **Ketua Jurusan Matematika**

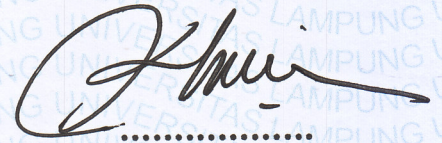
Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

MENGESAHKAN

1. Tim Penguji

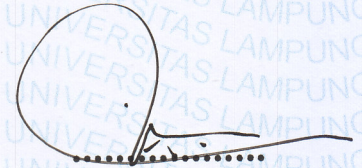
Ketua

: Dr. Khoirin Nisa, M.Si.



Sekretaris

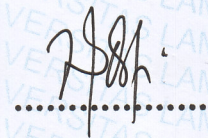
: Subian Saidi, S.Si., M.Si.



Penguji

Bukan Pembimbing

: Widiarti, S.Si., M.Si.

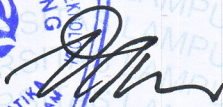


2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Supto Dwi Yuwono, S.Si., M.T.

NIP. 19740705 200003 1 001



Tanggal Lulus Ujian Skripsi: 20 Agustus 2021

PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Epmi Annisa Nur Fajrina
Nomor Pokok Mahasiswa : 17570310006
Jurusan : Matematika

Dengan ini menyatakan bahwa skripsi saya yang berjudul “**ANALISIS *CLUSTER ROBUST* MENGGUNAKAN METODE *K-MEDOIDS* PADA DATA MENGANDUNG PENCILAN**” adalah hasil pekerjaan saya sendiri. Semua hasil tulisan dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila dikemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau telah dibuat orang lain, maka saya bersedia menerima sanksi sesuai ketentuan akademik yang berlaku.

Bandar Lampung, 23 September 2021

Penulis



Epmi Annisa Nur Fajrina
NPM. 1757031006

RIWAYAT HIDUP

Penulis bernama lengkap Epmi Annisa Nur Fajrina. Lahir di Sukanegara pada tanggal 30 Oktober 1999, merupakan anak pertama dari tiga bersaudara, pasangan Bapak Eko Prayitno dan Ibu Ngatemi serta kakak dari M.Ilham Arief Himawan dan Epmi Zulfa Sandrina Ariedhani.

Penulis mengawali pendidikan kanak-kanak di TK ABA Sukanegara pada tahun 2003-2005. Kemudian penulis melanjutkan pendidikan Sekolah Dasar di SD Negeri 1 Sukanegara pada tahun 2005-2011. Selanjutnya penulis melanjutkan pendidikan Sekolah Menengah Pertama di SMP Negeri 1 Bangunrejo pada tahun 2011-2014 dan melanjutkan pendidikan Sekolah Menengah Atas di SMA Negeri 1 Kalirejo pada tahun 2014-2017.

Pada tahun 2017 penulis melanjutkan pendidikan Strata Satu (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung. Selama menjadi mahasiswa penulis aktif di beberapa organisasi yaitu Generasi Muda Himatika (GEMATIKA) 2017, Korps Muda BEM U KBM UNILA 2017, Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA Unila sebagai Anggota Biro Dana dan Usaha periode 2018, anggota staff kementerian Dalam Negeri BEM U KBM Unila tahun 2018, dan anggota staff komisi III DPM U KBM Unila pada tahun 2019.

Pada bulan Januari hingga Februari 2020 penulis melaksanakan Kerja Praktik (KP) di Badan Pusat Statistik Provinsi Lampung sebagai bentuk pengembangan diri serta menerapkan ilmu yang telah didapat selama perkuliahan. Selanjutnya pada bulan Juli hingga Agustus 2020 penulis melaksanakan Kuliah Kerja Nyata (KKN) Periode II di Desa Kalirejo, Kecamatan Kalirejo, Kabupaten Lampung Tengah sebagai bentuk pengabdian kepada masyarakat.

KATA INSPIRASI

“Barang siapa yang menempuh jalan untuk mencari ilmu, maka Allah akan memudahkan bagimu jalan menuju surga.”

(HR. Muslim)

“Wahai-orang yang beriman! Mohonlah pertolongan (kepada Allah) dengan sabar dan sholat.”

(QS. Al-Baqarah :153)

“Allah tidak akan membebani seseorang melainkan sesuai dengan kesanggupannya.”

(QS. Al-Baqarah :286)

“Berdoalah kepada-Ku, niscaya akan Aku perkenankan bagimu.”

(QS. Ghafir: 60)

“Optimism is the faith that leads to achievement. Nothing can be done without hope and confidence.”

(Hellen Keller)

PERSEMBAHAN

Puji dan syukur kepada Allah SWT yang telah memberikan rahmat, hidayah, dan karunia-Nya kepada ku.

Dengan penuh rasa cinta dan sayang serta mengharap keridhoan Allah SWT kupersembahkan karya sederhana ini kepada:

Bapak, Ibu, Adik, dan keluarga Besarku Tersayang

Yang senantiasa mencurahkan kasih sayang, doa, dan dukungan tiada henti. Terima kasih sudah menjadi rumah dan tempat terbaik untuk pulang dari segala keluh kesah atas hal-hal pelik yang datang.

Dosen Pembimbing dan Penguji

Yang telah meluangkan waktu untuk memberikan bimbingan, kritik, saran, serta ilmu yang sangat berharga kepada penulis.

Sahabat-Sahabatku

Yang selalu memberikan semangat, dukungan, doa, serta selalu membantu penulis selama masa perkuliahan.

Almamater Tercinta Universitas Lampung

SANWACANA

Alhamdulillah puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayah serta karunia-Nya kepada penulis sehingga skripsi dengan judul “**Analisis Cluster Robust Menggunakan Metode K-Medoids pada Data Mengandung Pencilan**” dapat terselesaikan dengan baik.

Terselesaikannya skripsi ini tidak lepas dari dukungan, bimbingan, saran, serta doa dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis mengucapkan terima kasih kepada:

1. Ibu Dr. Khoirin Nisa, M.Si., selaku Dosen Pembimbing I dan Dosen Pembimbing Akademik yang telah membimbing penulis, menyumbangkan ilmu, memberikan motivasi dan arahan, serta kesediaan waktu yang diberikan baik selama perkuliahan maupun penelitian.
2. Bapak Subian Saidi, S.Si., M.Si., selaku Dosen Pembimbing II yang telah membimbing penulis, menyumbangkan ilmu, memberikan motivasi dan arahan, serta kesediaan waktu yang diberikan selama penelitian.
3. Ibu Widiarti, S.Si., M.Si., selaku Dosen Penguji atas kesediaannya untuk menguji, memberikan kritik, dan saran yang membangun dalam menyelesaikan skripsi ini.
4. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

6. Seluruh dosen, staff, dan karyawan Fakultas Matematika dan Ilmu Pengetahuan Alam khususnya Jurusan Matematika yang telah memberikan ilmu dan bantuan yang berguna bagi penulis.
7. Bapak, Ibu, adik-adik yaitu Arief dan Zulfa, serta seluruh keluarga besar Prayitno dan keluarga besar Amat Kusni yang selalu menjadi penyemangat tersendiri bagi penulis dan tiada henti memberikan kasih sayang, dukungan, motivasi, serta doa kepada penulis.
8. Sahabat terbaik penulis, Indah, Dewi, Dhea, Yustika, Nita, Inas, Atina yang selalu siap mendengarkan keluh kesah , memberi semangat dan keceriaan serta selalu membantu penulis dalam hal apapun.
9. Yunda Siska dan bang Agung, yang selalu membantu memberikan solusi atas kendala yang dihadapi selama menyelesaikan skripsi.
10. Teman-teman Korps Muda BEM U KBM UNILA periode 2017, serta keluarga besar BEM U KBM UNILA periode 2017 yang telah memberikan banyak pembelajaran dan pengalaman kepada Penulis.
11. Anggota Staff Kementerian Dalam Negeri BEM U KBM UNILA periode 2018, serta keluarga besar BEM U KBM UNILA periode 2018 yang telah memberikan banyak pembelajaran dan pengalaman kepada Penulis.
12. Anggota Biro Dana dan Usaha Himatika periode 2018, serta keluarga besar HIMATIKA periode 2018 yang telah memberikan banyak pembelajaran dan pengalaman kepada Penulis.
13. Anggota Staff Komisi III DPM U KBM UNILA periode 2019, serta keluarga besar DPM U KBM UNILA periode 2019 yang telah memberikan banyak pembelajaran dan pengalaman kepada Penulis.
14. Teman-teman “Lambetika C” atas kebersamaan dan canda tawanya selama masa perkuliahan.
15. Teman-teman seperbimbingan sebagai tempat saling bertukar pikiran, memberikan saran, dan saling meberikan semangat selama proses penyelesaian skripsi.
16. Teman-teman Jurusan Matematika angkatan 2017 atas kebersamaannya.

17. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih terdapat kekurangan baik isi maupun susunannya. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun agar dapat menjadi bahan perbaikan untuk kedepannya.

Bandar Lampung, 23 September 2021

Penulis

Epmi Annisa Nur Fajrina

DAFTAR ISI

	Halaman
DAFTAR TABEL	vii
DAFTAR GAMBAR.....	viii
I. PENDAHULUAN	
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	4
1.3 Manfaat Penelitian	4
II. TINJAUAN PUSTAKA	
2.1 Analisis Klaster.....	5
2.1.1 Asumsi Analisis Klaster.....	5
2.1.2 Ukuran Kemiripan Objek (jarak)	6
2.2 Pencilan.....	7
2.3 Metode Hierarki	8
2.4 Metode Non Hierarki	10
2.4.1 Metode Partisi	11
2.4.1.1 Metode <i>K-Medoids</i>	11
2.5 Validasi Hasil Klaster	12
2.5.1 <i>Sillhouette Index</i>	13
2.5.2 <i>Dunn Index</i>	14
III. METODOLOGI PENELITIAN	
3.1 Waktu dan Tempat Penelitian	15
3.2 Data Penelitian.....	15
3.3 Metode Penelitian	16

IV. HASIL DAN PEMBAHASAN

4.1	Membangkitkan Data	19
4.2	Uji Asumsi Analisis Klater.....	20
4.2.1	Uji Asumsi Sampel Representatif dengan Pendeteksian Pencilan	20
4.2.2	Uji Asumsi Tidak Terjadi Multikolinearitas	22
4.3	Analisis Klaster Menggunakan Metode <i>K-Medoids</i>	23
4.4	Menghitung Nilai Indeks Validitas.....	28
4.4.1	<i>Sillhouette Index</i>	28
4.4.2	<i>Dunn Index</i>	29
4.5	Proses Pengulangan Sebanyak 1000 Kali.....	29
4.5.1	Nilai <i>sillhouette index</i> dan <i>Dunn index</i> hasil analisis klaster <i>K-medoids</i> pada data yang dibangkitkan sebanyak 1000 kali pengulangan untuk $n = 30$	30
4.5.2	Nilai <i>sillhouette index</i> dan <i>Dunn index</i> hasil analisis klaster <i>K-medoids</i> pada data yang dibangkitkan sebanyak 1000 kali pengulangan untuk $n = 50$	31
4.5.3	Nilai <i>sillhouette index</i> dan <i>Dunn index</i> hasil analisis klaster <i>K-medoids</i> pada data yang dibangkitkan sebanyak 1000 kali pengulangan untuk $n = 70$	32
4.5.4	Nilai <i>sillhouette index</i> dan <i>Dunn index</i> hasil analisis klaster <i>K-medoids</i> pada data yang dibangkitkan sebanyak 1000 kali pengulangan untuk $n = 100$	33
4.6	Analisis Hasil	34

V. KESIMPULAN

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

Tabel	Halaman
1. Contoh data simulasi untuk $n = 30$ dengan proporsi pencilan 5%	19
2. Nilai jarak kuadrat Mahalanobis masing-masing objek.....	21
3. Nilai VIF dari masing-masing variabel pada data contoh.....	22
4. Objek yang dipilih sebagai <i>medoid</i> awal.....	23
5. Hasil perhitungan jarak setiap objek dengan <i>medoid</i> awal	24
6. Objek yang dipilih sebagai <i>medoid</i> baru	25
7. Hasil perhitungan jarak setiap objek dengan <i>medoid</i> baru.....	25
8. Hasil klusterisasi metode <i>K-medoids</i>	27
9. Nilai <i>sillhouette_width</i> setiap objek pada data contoh.....	28
10. Nilai <i>sillhouette index</i> dan <i>Dunn index</i> hasil analisis kluster <i>K-medoids</i> pada data yang dibangkitkan sebanyak 1000 kali pengulangan untuk $n = 30$	30
11. Nilai <i>sillhouette index</i> dan <i>Dunn index</i> hasil analisis kluster <i>K-medoids</i> pada data yang dibangkitkan sebanyak 1000 kali pengulangan untuk $n = 50$	31
12. Nilai <i>sillhouette index</i> dan <i>Dunn index</i> hasil analisis kluster <i>K-medoids</i> pada data yang dibangkitkan sebanyak 1000 kali pengulangan untuk $n = 70$	32

13. Nilai <i>sillhouette index</i> dan <i>Dunn index</i> hasil analisis klaster <i>K-medoids</i> pada data yang dibangkitkan sebanyak 1000 kali pengulangan untuk $n = 100$	33
14. Rata-rata nilai <i>sillhouette index</i> dan <i>Dunn index</i> hasil analisis klaster <i>K-medoids</i> pada data yang dibangkitkan sebanyak 1000 kali pengulangan	34
15. Rata-rata nilai <i>sillhouette index</i> dan <i>Dunn index</i> berdasarkan jumlah objek (n)	36
16. Rata-rata nilai <i>sillhouette index</i> dan <i>Dunn index</i> berdasarkan jumlah klaster (k).....	37
17. Rata-rata nilai <i>sillhouette index</i> dan <i>Dunn index</i> berdasarkan proporsi pencilan (p).....	38

DAFTAR GAMBAR

Gambar	Halaman
1. <i>Flowchart</i> penelitian	18
2. Boxplot data contoh	22
3. Plot hasil kluster metode <i>K-medoids</i> pada data contoh	28
4. Plot rata-rata nilai <i>sillhouette index</i> dan <i>Dunn index</i> berdasarkan jumlah objek (n)	36
5. Plot rata-rata nilai <i>sillhouette index</i> dan <i>Dunn index</i> berdasarkan jumlah kluster (k)	37
6. Plot rata-rata nilai <i>sillhouette index</i> dan <i>Dunn index</i> berdasarkan proporsi pencilan (p)	38

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis kluster merupakan suatu teknik analisis multivariat yang berguna untuk mengelompokkan data observasi ataupun variabel-variabel ke dalam kluster sedemikian rupa sehingga masing-masing kluster bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran. Data mengenai ukuran kesamaan tersebut dapat dianalisis dengan analisis kluster sehingga dapat ditentukan siapa yang masuk kluster mana (Gudono, 2011).

Analisis kluster bertujuan menempatkan sekumpulan objek ke dalam dua atau lebih kluster berdasarkan kesamaan objek atas dasar berbagai karakteristik. Banyak objek yang dapat dikelompokkan dengan analisis kluster di antaranya adalah produk (barang dan jasa), manusia (responden, konsumen), dan benda (Supranto, 2004).

Analisis kluster dibagi menjadi dua metode yaitu hierarki (*hierarchical clustering methods*) dan metode non hierarki (*non hierarchical clustering methods*). Perbedaan antara kedua metode tersebut terletak pada penentuan jumlah kluster yang akan dihasilkan. Metode hierarki digunakan apabila jumlah kluster yang diinginkan tidak diketahui, sedangkan metode non hierarki digunakan apabila jumlah kelompok yang diinginkan telah ditentukan sebelumnya. Metode non hierarki yang sering digunakan di antaranya metode *K-means*.

Ada dua asumsi yang harus dipenuhi pada analisis kluster, yaitu sampel representatif dan non multikolinearitas antara tiap variabel (Hair, *et al.*, 2014). Sampel yang digunakan dalam analisis kluster harus dapat mewakili populasi yang ingin dijelaskan agar sampel representatif.

Pada kenyataannya, tidak semua data memenuhi kedua asumsi di atas khususnya untuk asumsi sampel representatif karena adanya masalah pencilan. Pencilan merupakan pengamatan yang tidak mengikuti sebagian besar pola dan terletak jauh dari pusat data. Adanya pencilan ini akan mempengaruhi kesimpulan dari analisis data. Namun, pencilan tidak dapat dibuang begitu saja, karena adakalanya pencilan memberikan informasi yang tidak dapat diberikan oleh data pengamatan lainnya. Untuk mengatasi masalah pencilan tersebut, maka diperlukan metode *robust*. Berbagai kajian mengenai analisis kluster *robust* telah dilakukan oleh para peneliti diantaranya Rifa, *et al.* (2020), Larasati, *et al.* (2021), Sari & Saputro (2021). Terdapat beberapa metode *robust* dalam analisis kluster non hierarki salah satunya yaitu *K-Medoids*.

Metode *K-Medoids* atau sering disebut juga PAM (*Partitioning Around Medoids*) diperkenalkan oleh Kaufman dan Rousseeuw (1990). Metode ini termasuk dalam teknik penyekatan (*partition*) yang membagi atau memisahkan objek ke-*k* kelompok yang berbeda sehingga objek yang memiliki karakteristik yang sama masuk ke dalam satu kluster yang sama dan objek yang mempunyai karakteristik berbeda masuk ke dalam kluster yang lain.

Algoritma *K-Medoids* muncul sebagai penanggulangan kelemahan Algoritma *K-Means* yang sensitif terhadap pencilan disebabkan sebuah objek dengan sebuah nilai yang besar memungkinkan menyimpang dari sebaran data.

Beberapa penelitian mengenai metode *K-medoids* dalam klastering di antaranya penelitian yang dilakukan oleh Flowrensia (2010) pada kasus klasterisasi karakteristik tanaman bunga iris menghasilkan metode *K-medoids* lebih baik dibandingkan metode *K-means* pada data yang mengandung pencilan pada kondisi proporsi sebesar 5%.

Penelitian oleh Milla, dkk (2019) tentang metode *K-medoids clustering* dengan validasi *sillhouette index* dan *C-index* pada kasus Jumlah Kriminalitas kabupaten/kota di Jawa Tengah Tahun 2018.

Berdasarkan uraian tersebut, peneliti tertarik untuk melakukan penelitian tentang analisis klaster *robust* metode *K-medoids* menggunakan simulasi pada data yang mengandung pencilan. Efektivitas metode tersebut ditinjau berdasarkan nilai validasi *sillhouette index* dan *Dunn index* yang dihasilkan.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini sebagai berikut.

1. Mengkaji analisis kluster *robust* dengan metode *K-medoids* melalui data simulasi dengan beberapa persentase nilai pencilan.
2. Melihat keefektifan metode *K-medoids* untuk mengatasi pencilan ditinjau dari nilai validasi *sillhouette index* dan *Dunn index*.

1.3 Manfaat Penelitian

Adapun manfaat yang diperoleh dari penelitian ini adalah sebagai berikut.

1. Memberikan informasi mengenai keefektifan metode *K-medoids* pada analisis kluster *robust* dalam mengatasi pencilan.
2. Memberikan metode alternatif bagi para peneliti dalam melakukan analisis klaster non hierarki apabila dihadapkan pada data yang mengandung pencilan.

II. TINJAUAN PUSTAKA

2.1 Analisis Klaster

Analisis klaster adalah teknik pengklasteran untuk mengelompokkan objek berdasarkan karakteristik yang dimiliki oleh objek tersebut. Objek diklasifikasikan ke dalam satu atau lebih klaster sehingga objek-objek yang berada di dalam klaster akan mempunyai kemiripan atau kesamaan karakter (Hair, *et al.*, 2014).

Analisis klaster merupakan suatu teknik analisis multivariat yang bertujuan untuk mengklasterkan data observasi ataupun variabel-variabel ke dalam klaster sedemikian rupa sehingga masing-masing klaster bersifat homogen sesuai dengan faktor yang digunakan untuk melakukan pengklasteran. Data mengenai ukuran kesamaan tersebut dapat dianalisis dengan analisis klaster sehingga dapat ditentukan siapa yang masuk klaster mana (Gudono, 2011).

2.1.1 Asumsi Analisis Klaster

Analisis klaster bukanlah teknik statistik inferensia di mana parameter dari sampel dapat menilai dan mewakili suatu populasi. Analisis klaster adalah metode untuk mengukur karakteristik struktural dari serangkaian pengamatan (Hair, *et al.*, 2014). Pada analisis klaster asumsi seperti normalitas, linearitas, dan homoskedastisitas tidak banyak berpengaruh.

Menurut Widarjono (2010), ada dua asumsi yang harus dipenuhi pada analisis klaster, yaitu:

1. Sampel representatif

Sampel yang digunakan dalam analisis klaster harus dapat mewakili populasi yang ingin dijelaskan. Pencilan merupakan salah satu masalah yang perlu diperhatikan, karena dengan adanya pencilan artinya terdapat sampel yang tidak mewakili populasi. Oleh karena itu, diperlukan metode yang *robust* terhadap pencilan agar sampel representatif.

2. Tidak terjadi multikolinearitas.

Multikolinieritas adalah suatu peristiwa terjadinya korelasi yang kuat antara dua atau lebih variabel. Multikolinearitas juga merupakan masalah yang perlu diperhatikan dalam analisis klaster, karena dapat mempengaruhi hasil pengklasteran akhir. Multikolinearitas dapat dilihat dari nilai *Variance Inflation Factor* (VIF). Jika nilai VIF melebihi angka 10 maka dapat disimpulkan ada multikolinearitas.

2.1.2 Ukuran Kemiripan Objek (jarak)

Klasterisasi dilakukan berdasarkan kemiripan antar objek. Kemiripan diperoleh dengan meminimalkan jarak antar objek dalam klaster dan memaksimalkan jarak antar klaster. Semakin kecil besaran jarak suatu objek terhadap objek lain, maka semakin besar kemiripan objek tersebut, sehingga objek tersebut akan dimasukkan dalam klaster yang sama (Usman & Nurdin, 2013). Ukuran jarak yang biasa digunakan dalam analisis klaster yaitu jarak euclidean. Rumus jarak euclidean dinyatakan sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (2.1)$$

dengan:

d_{ij} : jarak antara objek ke- i dan obyek ke- j

n : jumlah variabel kluster

x_{ik} : data dari objek ke- i pada variabel ke- k

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

2.2 Pencilan

Dalam analisis kluster pencilan merupakan suatu masalah yang penting karena membuat tidak terpenuhinya salah satu asumsi. Pencilan dapat diartikan sebagai suatu keganjilan karena jauh lebih kecil atau jauh lebih besar daripada sebagian besar pengamatan dalam populasi. Pencilan digambarkan sebagai observasi yang tidak mewakili populasi umum. Adanya pencilan dapat mengakibatkan kurang tepatnya hasil analisis yang diperoleh dan tidak mewakili keadaan populasi (Hair, *et al.*, 2014).

Salah satu cara yang sering digunakan untuk mendeteksi adanya pencilan dalam data pengamatan adalah menggunakan *boxplot*. *Boxplot* adalah gambaran secara grafis, berdasarkan kuartil, yang membantu kita dalam menggambarkan sekumpulan data (Lind, *et al.*, 2007). Selain itu, pendeteksian pencilan juga dapat dilakukan dengan cara menghitung kuadrat jarak Mahalanobis d_{ij}^2 dari masing-masing pengamatan. Jarak Mahalanobis antar dua objek dinyatakan dalam bentuk vektor dan matriks. Jarak antara objek \mathbf{x}_i dan \mathbf{x}_j dinyatakan sebagai berikut:

$$d_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \mathbf{x}_j) \quad (2.2)$$

dengan \mathbf{x}_i dan \mathbf{x}_j adalah vektor untuk objek ke- i dan ke- j . $\boldsymbol{\Sigma}$ merupakan matriks varian kovarian (Sharma, 1996). Suatu data dikatakan bebas dari pencilan apabila $d_{ij}^2 < \chi_{\alpha;p}^2$ dengan α merupakan taraf signifikansi dan p banyaknya variabel (Bowo, dkk. 2013).

Walaupun memiliki perilaku yang berbeda dengan mayoritas data yang lain dan sering dianggap *noise*, tetapi pencilan sering kali mengandung informasi yang sangat berguna. Tidak semua data yang mengandung pencilan bisa ditransformasi karena kasus data yang berbeda-beda. Akan tetapi, dengan menggunakan metode yang *robust*, pencilan tidak berpengaruh secara signifikan.

2.3 Metode Hierarki

Prosedur pembentukan kluster terbagi menjadi 2, yaitu hierarki dan non-hierarki. Pembentukan kluster hierarki mempunyai sifat sebagai pengembangan suatu hierarki atau struktur mirip pohon bercabang. Metode hierarki merupakan metode pengklasteran dengan jumlah kluster yang akan dibuat belum diketahui. Metode ini terbagi dalam 2 teknik yaitu penggabungan berurutan (*agglomerative*) atau pembagian berurutan (*divisive*).

Metode hierarki (*hierarchical methods*) merupakan salah satu teknik klasterisasi dengan cara mengelompokkan dua atau lebih objek yang mempunyai nilai kemiripan yang paling dekat. Selanjutnya hasil pengelompokkan pertama dikelompokkan lagi dengan objek lain yang mempunyai nilai kemiripan kedua. Demikian seterusnya, sehingga akan membentuk konstruksi hierarki atau berdasarkan tingkatan tertentu seperti struktur pohon (struktur pertandingan). Dendogram merupakan diagram yang digunakan untuk menggambarkan struktur dari hasil pengklasteran tersebut (Simamora, 2005).

Metode-metode analisis kluster hierarki yang sering digunakan adalah sebagai berikut.

1. Pautan tunggal (*single linkage*).

Pautan tunggal (*single linkage*) merupakan prosedur pengklasteran berdasarkan jarak terkecil antar objek. Algoritma pengklasteran *single linkage* diawali dengan memilih jarak terkecil dalam matriks $\mathbf{D} = \{\mathbf{d}_{ij}\}$ kemudian menggabungkan objek yang bersesuaian misalnya U dan V untuk memperoleh kluster (UV) . Langkah berikutnya adalah mencari jarak antara (UV) dengan kluster lainnya, misalnya W sehingga dapat dituliskan sebagai berikut:

$$\mathbf{d}_{(UV)W} = \min(\mathbf{d}_{UW}, \mathbf{d}_{VW}) \quad (2.3)$$

dengan \mathbf{d}_{UW} adalah jarak terdekat dari kluster U dan W serta \mathbf{d}_{VW} adalah jarak terdekat dari kluster V dan W .

2. Pautan penuh (*complete linkage*)

Pautan penuh (*complete linkage*) merupakan prosedur pengklasteran berdasarkan jarak terbesar antar objek. Algoritma pengklasteran *complete linkage* diawali dengan memilih jarak terbesar dalam matriks $\mathbf{D} = \{\mathbf{d}_{ij}\}$ kemudian menggabungkan objek yang bersesuaian misalnya U dan V untuk memperoleh kluster (UV) . Langkah berikutnya adalah mencari jarak antara (UV) dengan kluster lainnya, misalnya W sehingga dapat dituliskan sebagai berikut:

$$\mathbf{d}_{(UV)W} = \max(\mathbf{d}_{UW}, \mathbf{d}_{VW}) \quad (2.4)$$

3. Metode *centroid linkage*.

Pada metode *centroid linkage*, jarak antara dua kluster adalah jarak antar *centroid* kluster tersebut. *Centroid linkage* merupakan nilai tengah observasi pada variabel dalam suatu set variabel kluster. Keuntungan metode ini adalah pencilan hanya sedikit berpengaruh jika dibandingkan dengan metode lain.

4. Pautan rata-rata (*average linkage*).

Pautan rata-rata (*average linkage*) merupakan prosedur pengklasteran berdasarkan rata-rata antar objek. Algoritma *average linkage* diawali dengan mendefinisikan matriks $\mathbf{D} = (d_{ij})$ untuk memperoleh objek yang paling dekat, misalnya objek U dan V , kemudian objek ini digabung ke dalam bentuk kluster (UV) dan selanjutnya jarak antara (UV) dengan kluster lainnya W , sehingga dapat dituliskan sebagai berikut:

$$d_{(UV)W} = \frac{d_{(UW)} + d_{(VW)}}{n_{(UV)}n_W} \quad (2.5)$$

dengan $n_{(UV)}$ adalah banyaknya anggota kluster (UV) dan n_W adalah banyaknya anggota dalam cluster W .

5. Metode *Ward*.

Pada metode *Ward*, jarak antara dua kluster adalah jumlah kuadrat antara dua kluster untuk seluruh variabel. Metode ini cenderung digunakan untuk mengkombinasi kluster-kluster dengan jumlah kecil.

2.4 Metode Non Hierarki

Metode non hierarki digunakan untuk pengklasteran objek dimana banyaknya kluster yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur pengklasteran. Metode ini dapat diterapkan pada data yang lebih besar dibandingkan metode hierarki. Selain itu, metode non hierarki hanya memiliki sedikit kelemahan pada data mengandung pencilan, ukuran jarak yang digunakan, dan variabel yang tidak relevan (Simamora, 2005). Jenis-jenis metode non hierarki diantaranya *sequential threshold*, *parallel threshold*, dan *optimizing partitioning* (metode partisi). Pada penelitian ini akan dikaji lebih dalam mengenai metode non hierarki *optimizing partitioning* (metode partisi) pada data yang mengandung pencilan.

2.4.1 Metode Partisi

Metode partisi merupakan suatu metode yang melakukan optimasi pada penempatan objek yang ditukar untuk kluster lainnya dengan pertimbangan optimasi. Metode partisi ini didasarkan pada penentuan awal jumlah kluster, selanjutnya merelokasi objek secara iteratif untuk menemukan kembali kluster-kluster yang berada dalam satu titik, kemudian menempatkan kembali objek-objek ke dalam kluster yang lebih dekat (Sarwono, 2013). Ketika mempartisi satu set objek menjadi k kluster, tujuan utamanya adalah untuk menemukan kluster berdasarkan objek-objek yang menunjukkan tingkat kesamaan yang tinggi. Pada metode partisi terdapat beberapa metode yang sering digunakan antara lain metode *K-Medoids* dan metode *K-Means*.

2.4.1.1 Metode *K-Medoids*

Metode *K-medoids* atau sering disebut juga PAM (*Partitioning Around Medoids*) merupakan salah satu metode partisi yang digunakan untuk mengelompokkan sekumpulan n objek ke dalam k kluster. Metode pengklasteran ini menggunakan *medoid* sebagai pusat klasternya. *Medoid* merupakan objek yang letaknya terpusat di dalam suatu kluster. Keunggulan dari metode ini adalah kekuatannya, yaitu tidak terpengaruh terhadap adanya data pencilan ataupun data ekstrem (Bhat, 2014).

Perbedaan antara metode *K-Medoids* dengan metode *K-Means* yaitu metode *K-Medoids* menggunakan objek *medoid* sebagai perwakilan pusat kluster untuk setiap kluster, sementara metode *K-Means* membutuhkan nilai rata-rata (*mean*) sebagai pusat kluster (Kaur, *et al.*, 2014).

Langkah-langkah metode *K-Medoids* adalah sebagai berikut:

1. Inisialisasikan pusat kluster sebanyak jumlah kluster (k).
2. Setiap data atau objek dialokasikan ke kluster terdekat menggunakan persamaan ukuran jarak euclidian dengan rumus pada persamaan 2.1
3. Pilih objek pada masing-masing kluster secara acak sebagai kandidat *medoid* baru.
4. Hitung jarak setiap objek yang terdapat pada masing-masing kluster dengan calon *medoid* baru.
5. Hitung total simpangan (S) dengan menghitung nilai total jarak baru – total jarak lama. Jika didapatkan $S < 0$, tukarlah objek dengan data kluster untuk membuat sekumpulan k objek baru sebagai *medoid*.
6. Ulangi langkah 3 sampai dengan 5 hingga tidak terjadi perubahan *medoid*, sehingga diperoleh kluster serta anggota kluster masing-masing.

2.5 Validasi Hasil Kluster

Validasi hasil analisis kluster dilakukan untuk memperoleh partisi yang paling sesuai dengan data. Validasi jumlah kluster dilakukan dengan menggunakan *cluster validity index* atau indeks validitas. Indeks validitas berfungsi mengukur derajat kekompakan dan separasi struktur data pada seluruh kluster dan menemukan jumlah kluster optimal yang kompak dan terpisah dari kluster yang lain (Wu & Yang, 2004). Pada penelitian ini digunakan validasi *sillhouette index* dan *Dunn index* sebagai alat ukur validitas dalam analisis kluster *K-medoids*.

2.5.1 *Sillhouette Index*

Metode validasi *sillhouette index* merupakan metode yang digunakan untuk melihat kualitas dan kekuatan dari klaster. *Sillhouette index* akan mengevaluasi penempatan setiap objek dalam setiap klaster dengan membandingkan jarak rata-rata objek dalam satu klaster dan jarak antara objek dengan klaster yang berbeda (Aini, *et al.*, 2014).

Tahapan perhitungan *sillhouette index* sebagai berikut:

1. Hitung rata-rata jarak objek dengan semua objek lain yang berada di dalam satu klaster dengan persamaan berikut.

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.6)$$

2. Hitung rata-rata jarak objek dengan semua objek lain yang berada pada klaster lain, kemudian ambil nilai paling minimum dengan persamaan :

$$b(i) = \min d(i, C); \quad d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (2.7)$$

3. Hitung nilai *sillhouette _width* untuk masing-masing objek dengan persamaan :

$$s(i) = \frac{b(i)-a(i)}{\max(a(i), b(i))} \quad (2.8)$$

3. Hitung nilai *sillhouette index* dengan persamaan :

$$SI = \frac{1}{n} \sum_{i=1}^n s(i) \quad (2.9)$$

Nilai hasil *sillhouette index* terletak pada kisaran nilai -1 hingga 1. Semakin nilai *sillhouette index* mendekati 1, maka semakin baik pengelompokan data dalam satu klaster.

2.5.2 *Dunn Index*

Dunn index adalah salah satu pengukur validitas kluster yang diajukan oleh J.C.Dunn. Ukuran validitas kluster ini berlandaskan pada fakta bahwa kluster yang terpisah itu biasanya memiliki jarak antar kluster yang besar dan diameter intra kluster yang kecil (Satono, dkk. 2015).

Dunn index dapat dituliskan sebagai berikut:

$$D = \min_{j=i+1, \dots, n_c} \left(\frac{d(c_i, c_j)}{\max_{k=1, \dots, n_c} (\text{diam}(c_k))} \right) \quad (2.10)$$

Dimana nilai $d(c_i, c_j)$ dan $\text{diam}(c_k)$ didefinisikan sebagai berikut:

$$d(c_i, c_j) = \min_{\substack{x \in c_i \\ y \in c_j}} (d(x, y)) \quad (2.11)$$

$$\text{diam}(c_k) = \max_{\substack{x \in c_k \\ y \in c_k}} (d(x, y)) \quad (2.12)$$

Dunn index memiliki rentang nilai dari nol sampai tak hingga. Jika nilai *Dunn index* semakin besar, maka hasil kluster akan semakin baik.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester genap tahun akademik 2020/2021 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Pada penelitian ini data yang digunakan adalah data simulasi hasil dari membangkitkan data yang mengandung pencilan dengan menggunakan *software* RStudio versi 4.0.3. Banyaknya pencilan data yang dibangkitkan terdiri atas 0%, 5%, 10%, dan 20% dari keseluruhan data bangkitan. Banyak objek (n) yang dibangkitkan terdiri atas 30, 50, 70, dan 100 data bangkitan.

3.3 Metode Penelitian

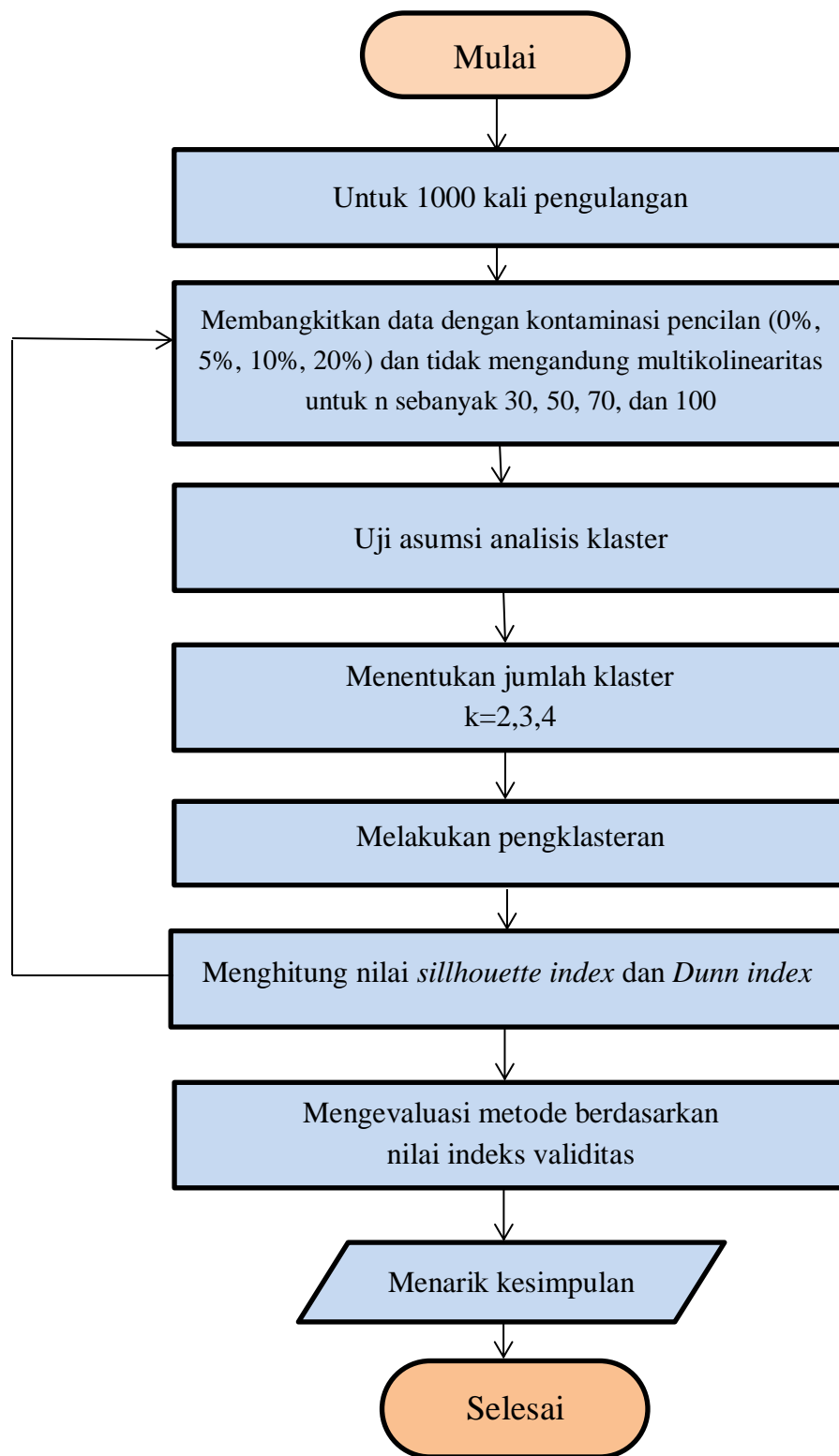
Penelitian ini dilakukan secara studi pustaka yaitu mempelajari buku-buku teks, jurnal serta akses internet yang menunjang proses penelitian. Adapun langkah-langkah penelitian yang dilakukan sebagai berikut:

1. Membangkitkan data penelitian.
 - a) Membangkitkan data berdistribusi normal dengan 5 variabel bebas. Masing-masing variabel bebas memiliki distribusi yaitu $X_1 \sim N(10,1)$, $X_2 \sim N(10,1)$, $X_3 \sim N(10,2)$, $X_4 \sim N(10,2)$ dan $X_5 \sim N(10,3)$. Kemudian menggabungkannya menjadi satu gugus data dan mencari μ dan matriks kovarian (Σ) dari data. Sehingga diperoleh

$$\mu = (10,10,10,10,10)$$
 dan

$$\Sigma = \begin{bmatrix} 1.003 & -0.162 & 0.219 & -0.142 & -0.799 \\ -0.162 & 0.710 & 0.339 & 0.200 & 0.073 \\ 0.219 & 0.339 & 4.151 & -0.029 & 0.386 \\ -0.142 & 0.200 & -0.029 & 2.708 & -0.524 \\ -0.799 & 0.073 & 0.386 & -0.524 & 9.670 \end{bmatrix} \quad (3.1)$$
 - b) Selanjutnya membangkitkan data berdistribusi normal multivariat dengan 5 variabel bebas untuk setiap n sehingga $X_i \sim \mathbf{N}_5(n, \mu, \Sigma)$ dengan $\mu = (10,10,10,10,10)$ dan Σ pada persamaan 3.1, dengan $i = 1,2,3,4,5$.
 - c) Kemudian membangkitkan data pencilan sebesar 5%, 10%, dan 20% yang memiliki distribusi normal (50,1).
 - d) Menggabungkan data beserta pencilannya ke dalam 1 gugus data dengan ketentuan data ke-1 sampai ke- p sebagai pencilan. Dengan $p =$ jumlah pencilan.
2. Melakukan uji asumsi analisis klaster.

3. Menentukan jumlah klaster yang akan dibentuk yaitu 2 klaster, 3 klaster, dan 4 klaster.
4. Melakukan pengklasteran dengan menggunakan metode *K-medoids*.
5. Menghitung dan mencatat nilai *sillhouette index* dan *Dunn index* untuk berbagai proporsi pencilan.
6. Mengulang langkah 1 (satu) sampai langkah 5 (lima) sebanyak 1000 (seribu) kali pengulangan.
7. Melakukan evaluasi keefektifan metode *K-medoids* pada berbagai proporsi pencilan berdasarkan nilai *sillhouette index* dan *Dunn index* yang diperoleh.
8. Menarik kesimpulan.



Gambar 1. Flowchart penelitian.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan yang telah dijelaskan pada Bab IV, maka dapat diambil kesimpulan sebagai berikut:

1. Analisis kluster metode *K-medoids* merupakan metode pengklasteran yang efektif digunakan pada data yang mengandung pencilan.
2. Berdasarkan rata-rata nilai *sillhouette index* dan *Dunn index* diperoleh hasil bahwa metode *K-medoids* efektif digunakan untuk pengklasteran pada data dengan jumlah objek kecil, jumlah kluster kecil, dan proporsi pencilan yang besar.

DAFTAR PUSTAKA

- Aini, F.N., Palgunadi, S., & Anggrainingsih, R. 2014. Clustering Business Recess Model Petri Net dengan Complete Linkage. *ITSMART*. **3**(2): 47-51.
- Bhat, A. 2014. K-Medoids Clustering Using Partitioning Around Medoids For Performing Face Recognition. *International Journal of Soft Computing, Mathematics and Control*. **3**(3): 1–12.
- Bowo, K., Hoyyi, A., & Mukid, M. 2013. Analisis Faktor-Faktor yang Mempengaruhi Keputusan Pembelian dan Kepuasan Konsumen pada Notebook Merek Acer (Studi Kasus Mahasiswa Universitas Diponegoro). *Jurnal Gaussian*, 29-38. Universitas Diponegoro, Semarang.
- Flowrensia, Y. 2010. *Perbandingan Penggerombolan K-Means dan K-Medoids Pada Data Yang Mengandung Pencilan*. Skripsi. Departemen Statistika Institut Pertanian Bogor, Bogor.
- Gudono. 2011. *Analisis Data Multivariat*. BPFE, Yogyakarta.
- Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. 2014. *Multivariate Data Analysis*. 7th Edition. Pearson Education Limited, England.
- Kaufman, L., & Rousseeuw, P.J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley, New York.

- Kaur, N.K., Kaur, U., & Singh, D. 2014. K-Medoids Clustering Algorithm – A Review. *International Journal of Computer Application and Technology (IJCAT)*. **1**:1841-2349.
- Larasati, S.D.A., Nisa, K. & Herawati, N. 2021. Robust Principal Component Trimmed Clustering of Indonesian Provinces Based on Human Development Index Indicators. *Journal of Physics: Conference Series*. **1751** (2021):1-8.
- Lind., Marchal., & Watchen. 2007. *Teknik-teknik Statistika dalam Bisnis dan Ekonomi Menggunakan Kelompok Data*. Edisi ke-13. Diterjemahkan oleh Chriswan Sungkono. Salemba Empat, Jakarta.
- Nahdliyah, M.A., Widiharih, T., & Prahutama, A. 2019. Metode *K-Medoids Clustering* Dengan Validasi *Silhouette Index* Dan *C-Index* . *Jurnal Gaussian*. **1**: 161-170.
- Rifa, I.H, Pratiwi, H., & Respatiwan, R. 2020. Clustering of Earthquake Risk in Indonesia Using K-medoids and K-means Algorithms. *MEDIA STATISTIKA*. **13**:194-205.
- Sari, I.A. & Saputro, D.R.S. 2021. Algoritma *Quick Robust Clustering Using Links* (QROCK) untuk *Clustering* Data Kategorik. *PRISMA*. **4**: 640-644.
- Sarwono, J. 2013. *Statistik Multivariat Aplikasi untuk Riset Skripsi*. C.V ANDI OFFSET, Yogyakarta.
- Satono, B.D., Khotimah, B.K., & Muhammad, A. 2015. Pengelompokan Tingkat Kesehatan Masyarakat Menggunakan *Shelf Organizing Maps* dengan *Cluster Validation Idb* dan *I-Dunn*. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*.
- Sharma, S. 1996. *Applied Multivariate Techniques*. A John Wiley & Sons, Canada.
- Simamora, B. 2005. *Analisis Multivariat Pemasaran*. PT Gramedia Pustaka Utama, Jakarta.

Supranto. 2004. *Analisis Multivariat: Arti dan Interpretasi*. Rineka Cipta, Jakarta.

Usman, H. & Nurdin, S. 2013. *Aplikasi Teknik Multivariate untuk Riset Pemasaran*. PT Raja Grafindo Persada, Jakarta.

Widarjono, A. 2010. *Analisis Statistika Multivariat Terapan*. UPP STIM YKPN, Yogyakarta.

Wu, K.L. & Yang, M.S. 2004. A Cluster Validity Index for Fuzzy Clustering. *Pattern Recognition Letters*. **26**(9): 1275-1291.