PENERAPAN METODE *TWO-STEP CLUSTERING* UNTUK DATA DENGAN VARIABEL CAMPURAN

(Skripsi)

Oleh

NURMA YUNITA

NPM 1717031089

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

ABSTRAK

PENERAPAN METODE *TWO STEP CLUSTERING* UNTUK DATA DENGAN VARIABEL CAMPURAN

Oleh

NURMA YUNITA

Pengklasteran adalah proses mengelompokkan objek ke dalam kelompk-kelompok yang memiliki kemiripan karakteristik. Metode pengklasteran yang sering digunakan adalah metode hierarki dan non hierarki. Kedua metode tersebut mensyaratkan jika peubah bersifat numerik. Apabila peubah bersifat campuran yaitu numerik dan kategorik maka digunakan *two step cluster.* Pada penelitian ini bertujuan menerapkan metode *two step cluster* dengan mengelompokkan kabupaten/kota di provinsi Jawa Barat menggunakan 6 peubah karakteristik demografi terdiri dari 2 peubah kategorik dan 4 peubah numerik. Berdasarkan hasil analisis diperoleh bahwa metode *two step cluster* mengelompokkan data kabupaten/kota di provinsi Jawa Barat berdasarkan peubah karakteristik demografi menjadi 2 *cluster.* Hasil 2 *cluster* ini merupakan hasil optimal berdasarkan nilai koefisien *silhouette* yaitu lebih dari 0,5 artinya *cluster* yang terbentuk memiliki struktur yang baik.

Kata Kunci: *Cluster, Two Step Cluster, Silhouette*.

ABSTRACT

APPLICATION OF THE TWO STEP CLUSTERING METHODS FOR DATA WITH MIX VARIABLES

By

NURMA YUNITA

Clustering is a process of classifying objects into groups that have characteristic similarities. The common clustering methods are hierarchical and non-hierarchical. The numerical variable is requirement for these methods. If the variable with mix of numerical and categorical, then two-step clustering is used. This study aims to apply the Two Step Cluster method by classifying regencies/cities in Jawa Barat Province. This analysis was used 6 demographics variables consist of two categorical variables and four numerical variables. Based on the analysis results, it is obtained that the two step clustering method classifies the regencies/cities data into 2 clusters. The result of these 2 cluster is optimal based on the silhouette coefficient value which is more than 0,5, it means that the cluster formed has a strong structure.

Keywords: Cluster, Two Step Cluster, Silhouette.

PENERAPAN METODE *TWO-STEP CLUSTERING* UNTUK DATA DENGAN VARIABEL CAMPURAN

Oleh

NURMA YUNITA

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2021

Judul Skripsi	: **PENERAPAN METODE *TWO-STEP CLUSTERING* UNTUK DATA DENGAN VARIABEL CAMPURAN**
Nama Mahasiswa	: Nurma Yunita
Nomor Pokok Mahasiswa	: 1717031089
Jurusan	: Matematika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Dr. Khoirin Nisa, S.Si., M.Si.	**Amanto, S.Si., M.Si.**
NIP 19740726 200003 2 001	NIP 19730314 200012 1 002

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Khoirin Nisa, S.Si., M.Si.** .....

Sekretaris : **Amanto, S.Si., M.Si.** .....

Penguji
Bukan Pembimbing : **Drs. Eri Setiawan, M.Si.** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Suripto Dwi Yuwono, M.T.
NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : **07 Desember 2021**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama Mahasiswa	: **Nurma Yunita**
Nomor Pokok Mahasiswa	: **1717031089**
Jurusan	: **Matematika**
Judul Skripsi	: **Penerapan Metode *Two Step Clustering* Untuk Data Dengan Variabel Campuran**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila dikemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 7 Desember 2021

Yang menyatakan

Nurma Yunita

RIWAYAT HIDUP

Penulis lahir di Desa Padang Cermin Kec. Way Khilau Kab. Pesawaran pada 7 Juni tahun 1999. Penulis merupakan anak pertama dari empat bersaudara, dari pasangan Bapak Saeruddin dan Ibu Zulaikho.

Riwayat pendidikan Sekolah Dasar di SD N 6 Pesawaran Kab. Pesawaran lulusan tahun 2011. Sekolah Menengah Pertama di MTs N 1 Pesawaran Kab. Pesawaran lulusan tahun 2014, dan Sekolah Menengah Atas di MAN 1 Pesawaran Kab. Pesawaran lulusan tahun 2017.

Penulis, melanjutkan jenjang pendidikan di Universitas Lampung jalur Beasiswa PMPAP tahun 2017 dan terdaftar sebagai mahasiswa jurusan Matematika FMIPA. Pada tahun 2020 Penulis melakukan Kuliah Kerja Nyata (KKN) di Desa Padang Cermin Kec. Way Khilau Kab. Pesawaran, dan Kuliah Praktik (KP) di Badan Pusat Statistik (BPS) Kota Bandar Lampung. Riwayat organisasi yakni Sains dan Teknologi (2018), Rohani Islam (2018) dan Gebyar Pelajar Lampung (2019).

KATA INSPIRASI

Ketahuilah...
"...bisa jadi kenyataan hari esok adalah impian kita hari ini..."
– Hasan Al Banna

"Apabila kita punya keinginan yang besar, kita juga harus punya keberanian yang sepadan untuk berjuang mewujudkannya."
– Alfi Alghazi

"Sesungguhnya Allah tidak akan mengubah keadaan suatu kaum hingga mereka merubah keadaan yang ada pada diri mereka sendiri."
– Q.S. Ar Rad :11

"Setiap muslim mempunyai senjata. Senjata yang tak tampak di hadapan mata tapi terasa jelas efeknya, yaitu doa."
– Alfi Alghazi

"Dan barang siapa yang bertaqwa kepada Allah, niscaya Allah menjadikan baginya kemudahan dalam urusannya."
– Q.S. At-Talaq : 4

PERSEMBAHAN

Puji dan syukur saya ucapkan kepada Allah SWT atas segala rahmat dan karunia-Nya. Tak lupa, selawat beserta salam senantiasa tercurahkan kepada baginda besar Nabi Muhammad SAW yang merupakan suri tauladan terbaik bagi seluruh umat.

Kupersembahkan sebuah karya sederhana ini kepada:

Kedua Orang Tua Tercinta

Terima kasih atas segala hal yang telah kalian berikan, baik dari segi materi, waktu luang, maupun doa yang tiada terhingga. Semoga kebaikan Bapak dan Ibu dibalas oleh Allah SWT dan keberkahan hidup selalu menyertai kalian.

Sahabat Terbaik

Terima kasih atas seluruh dukungan yang kalian berikan, semoga hubungan baik kita selalu terjaga tidak hanya di dunia, tetapi juga di akhirat.

Almamater Kebanggaan Universitas Lampung.

SANWACANA

Puji syukur penulis ucapkan kepada Allah SWT atas segala ridha, rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan tepat waktu. Skripsi dengan judul **"Penerapan Metode *Two Step Clustering* Untuk Data Dengan Variabel Campuran"** disusun untuk memenuhi salah satu syarat memperoleh gelar sarjana Matematika di Universitas Lampung.

Pada kesempatan kali ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Khoirin Nisa, M.Si. selaku pembimbing I yang telah memberikan masukan, saran serta bimbingan kepada penulis selama proses pembuatan skripsi ini hingga selesai.
2. Bapak Amanto, S.Si., M.Si. selaku pembimbing II serta pembimbing akademik yang telah memberikan bimbingan, arahan selama perkuliahan.
3. Bapak Drs. Eri Setiawan, S.Si. selaku dosen pembahas yang telah memberikan evaluasi, arahan serta saran hingga terselesaikannya 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. Suripto Dwi Yuwono, M.T. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Seluruh civitas akademik, dosen dan staf Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Ibu, Bapak dan Anggota Keluarga yang selalu memberikan dukungan dan doa terbaik agar penulis diberikan kelancaran serta kemudahan dalam menyelesaikan skripsi ini.

8. Teman-teman Terbaik; Reza, Eka dan Bila yang selalu memberikan dukungan moral dalam mengerjakan skripsi, pelajaran hidup dan nasihat.
9. Teman-teman Lambetika C jurusan Matematika yang telah membantu dan membersamai penulis selama perkuliahan.
10. Seluruh teman-teman mahasiswa jurusan matematika angkatan 2017.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat dibutuhkan, guna penyempurnakan skripsi ini.

Bandar Lampung, 7 Desember 2021
Penulis,

Nurma Yunita

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Gambar Halaman

DAFTAR TABEL

Tabel Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Menurut Ediyanto, dkk (2013), analisis *cluster* adalah teknik multivariat yang memiliki tujuan utama untuk mengklasifikasikan objek berdasarkan karakteristik yang dimiliki oleh objek tersebut. Pada analisis *cluster* dilakukan pengelompokan objek sehingga setiap objek yang saling berdekatan dan memiliki kesamaan dengan objek lainnya maka objek tersebut berada dalam *cluster* yang sama. *Cluster-cluster* yang terbentuk mempunyai homogenitas internal yang besar serta heterogenitas eksternal yang besar. Berbeda dengan metode multivariat yang lain, analisis ini tidak mengestimasi set peubah secara empiris sebaliknya yakni set peubah ditetapkan oleh pengamat itu sendiri.

Menurut Hair, *et al* (2019), analisis *cluster* biasanya melibatkan setidaknya tiga langkah. Langkah pertama adalah pengukuran terhadap beberapa bentuk kemiripan atau kumpulan antar objek untuk menentukan berapa banyak kelompok yang benar-benar ada dalam sampel. Langkah kedua adalah proses *clustering,* dimana objek dipartisi menjadi beberapa kelompok (*cluster*). Langkah terakhir adalah membuat profil peubah untuk menentukan komposisinya. Seringkali pembuatan profil ini dapat dilakukan dengan menerapkan diskriminan analisis terhadap kelompok yang diidentifikasi dengan teknik *cluster*.

Analisis *cluster* merupakan analisis multvariat yang termasuk dalam metode interdependensi, dimana peubah bebas x atau faktor penyebab memiliki kesamaan

dengan peubah terikat y atau respon. Untuk menentukan bahwa kedua objek dikatakan mirip dapat dilakukan dengan memperoleh matriks *proximity* yaitu matriks persegi dan simetris dengan jumlah objek yang sama pada baris dan kolom matriks, hal ini menunjukkan kemiripan atau ketidakmiripan antar objek (Nugroho, 2008).

Analisis *cluster* memiliki kelebihan salah satunya ialah dapat mengelompokkan data dalam jumlah yang cukup besar dan peubah yang banyak, serta memiliki kelemahan dengan peubah yang banyak kemungkinan terjadinya *error* menjadi lebih besar. Analisis *cluster* juga sangat banyak digunakan berbagai disiplin ilmu seperti kimia, biologi, ekonomi, sosial, kesehatan, teknis, bisnis dan bidang lainnya. Sehingga dapat disimpulkan bahwa analisis *cluster* memiliki begitu banyak kebermanfaatan (Rachmatin, 2014).

Menurut Hapsari, dkk (2020), pengklasifikasian objek pada metode analisis *cluster* umumnya hanya menggunakan peubah kontinu, oleh karena itu dikembangkan metode analisis untuk menyelesaikan peubah campuran yang disebut dengan analisis *two step cluster*. Mongi (2015) dalam penelitiannya mengungkapkan bahwa analisis *two step cluster* adalah metode pengklasteran yang dapat memberikan solusi untuk mengelompokkan suatu objek kedalam kelompok-kelompok yang memiliki kemiripan (homogen) dengan permasalahan pada skala pengukuran, objek yang diteliti berukuran cukup besar dengan peubah yang bebeda yaitu kategorik dan numerik sehingga hasil akhir untuk penyelesaian dari metode tersebut dapat diketahui *cluster* optimal yang terbentuk.

Peneliti tertarik untuk menerapkan metode *two step cluster* dalam pengelompokkan kabupaten/kota berdasarkan sosial dan kependudukan di wilayah provinsi Jawa Barat. Penelitian ini dilakukan sekiranya mampu dijadikan sebagai pertimbangan bagi pelaku kebijakan dalam menerapkan suatu program untuk mengatasi ketimpangan sosial di wilayah Jawa Barat. Hasil dari penelitian

ini yakni untuk melihat seberapa pengaruh suatu peubah pada masing-masing *cluster* yang terbentuk. Sehingga *cluster* yang terbentuk pada tiap-tiap wilayah dapat disesuaikan akan kebutuhannya dilihat dari topografi, demografi pada masing-masing kabupaten/kota yang berbeda-beda sehingga karakteristik dan kebutuhannya akan berbeda pula.

1.2 Tujuan Penelitian

Tujuan dari penelitian ini ialah untuk mengkaji dan menerapkan analisis *two step cluster* pada pengelompokkan kabupaten/kota berdasarkan sosial dan kependudukan di wilayah Provinsi Jawa Barat.

1.3 Manfaat penelitian

Manfaat dari penelitian ini adalah:

1. Dapat mengetahui hasil analisis *two step cluster* pada pengelompokkan kabupaten/kota berdasarkan sosial dan kependudukan di wilayah Jawa Barat.
2. Dapat dijadikan sebagai pertimbangan bagi pelaku kebijakan dalam menerapkan suatu program untuk mengatasi ketimpangan sosial di wilayah Jawa Barat melalui karakteristik suatu individu atau kelompok.
3. Penelitian ini mampu diharapkan menjadi bahan informasi untuk penelitian selanjutnya.

II. TINJAUAN PUSTAKA

2.1 Standarisasi/Pembakuan Data

Menurut Hair, *et al* (2019), pembakuan data adalah proses mengkonversi nilai masing-masing peubah awal menjadi nilai standar dengan rata-rata 0 dan standar deviasi 1 untuk menghilangkan bias yang disebabkan karena perbedaan skala dari beberapa peubah yang digunakan dalam analisis. Nilai standar untuk x_{ij} adalah:

$$Z_{ij} = \frac{x_{ij} - m_j}{s_j} \qquad (2.1)$$

dengan:

x_{ij} = nilai objek ke-i pada peubah ke-j

i =1,2,3, ... ,n

Z_{ij} = data x_{ij} yang sudah terstandarkan

S_j = simpangan baku dari peubah ke-j

m_j = rata-rata dari peubah ke-j

Menurut Sumertajaya, dkk (2007), jika peubah menggunakan satuan berbeda, peubah perlu distandarkan terlebih dahulu baru dilakukan analisis *cluster*.

2.2 Analisis *Cluster*

Analisis *cluster* adalah salah satu teknik multivariat yang digunakan untuk menganalisis data berdasarkan kriteria yang dimiliki. Analisis *cluster*

mengklasifikasikan objek ke dalam kelompok yang relatif homogen. Sehingga antara satu objek dengan objek lainnya yang berada dalam satu *cluster* cenderung mirip dan berbeda jauh dengan objek dari *cluster* lainnya. Hasil dari pengklasteran akan menunjukkan keragaman yang homogen di dalam *cluster* dan keragaman heterogen antar *cluster* yang terbentuk (Setiawan dan Pratiwi, 2019).

Menurut Mongi (2015), analisis *cluster* adalah analisis statistik dimana peubah ganda yang digunakan terdapat *n* buah objek yang memiliki *p* peubah yang akan dikelompokkan kedalam *k* kelompok. Objek yang terdapat dalam satu *cluster* mempunyai kemiripan yang lebih besar dibandingkan dengan objek yang terdapat dalam *cluster* lain.

Pengklasteran objek dilakukan dengan menggabungkan dua atau lebih objek membentuk suatu *cluster*, umumnya menggunakan suatu ukuran yang memiliki kemiripan atau ketidakmiripan. Apabila kemiripan dua objek semakin tinggi maka peluang untuk dikelompokkan dalam satu *cluster* semakin tinggi pula. Sebaliknya apabila semakin tidak mirip dua objek maka semakin rendah peluang untuk dikelompokkan dalam satu *cluster* (Lathifaturrahmah, 2014).

2.3 Ukuran Jarak

Ukuran jarak diperlukan untuk setiap pasang objek yang akan dikelompokkan. Ada beberapa metode pengukuran jarak antar dua objek, yaitu:

2.3.1 Jarak *Euclidean*

Jarak *Euclidean* merupakan jarak yang paling umum dan sering digunakan dalam analisis *cluster* apabila peubahnya berskala kontinu. Jarak *Euclidean* harus memenuhi asumsi jika peubah-peubah yang diamati tidak berkolerasi dan antar

peubah memiliki antar satuan yang sama. Pada metode ini, pengukuran jarak dilakukan dengan menghitung akar kuadrat dari penjumlahan kuadrat selisih dari nilai masing-masing peubah. Jarak *Euclidean* dapat didefinisikan sebagai berikut :

$$d_{i,j} = \sqrt{\sum_{k=1}^{p}\left(x_{ik} - x_{jk}\right)^2} \tag{2.2}$$

dengan:

$d_{i,j}$ = jarak antara objek *i* dengan objek *j*

x_{ik} = nilai objek *i* pada peubah ke-*k*

x_{jk} = nilai objek *j* pada peubah ke-*k*

p = banyaknya peubah yang diamati

(Mongi, 2015)

2.3.2 Jarak Manhattan (*City Block*/Minkowski)

Jarak mahhattan merupakan bentuk umum dari jarak *Euclidean*. Jarak manhattan digunakan jika peubah yang diamati berkorelasi atau tidak saling bebas. Pada metode ini, pengukuran jarak dilakukan dengan menghitung jumlah obsolut perbedaan objek untuk masing-masing peubah. Jarak manhattan dapat didefinisikan sebagi berikut:

$$d_{i,j} = \sum_{k=1}^{p}\left|x_{ik} - x_{jk}\right| \tag{2.3}$$

dengan:

$d_{i,j}$ = jarak antara objek *i* dengan objek *j*

x_{ik} = nilai objek *i* pada peubah ke-*k*

x_{jk} = nilai objek *j* pada peubah ke-*k*

p = banyaknya peubah yang diamati

(Mongi, 2015)

2.3.3 Jarak Mahalonobis

Jarak mahalonobis adalah jarak sangat diperlukan dalam menghilangkan atau mengatasi perbedaan skala pada masing-masing peubah. Jarak mahalonobis dapat didefinisikan sebagai berikut:

$$d_{i,j} = \sqrt{(x_{ik} - x_{jk})' S^{-1} (x_{ik} - x_{jk})} \tag{2.4}$$

dengan:

$d_{i,j}$ = jarak antara objek *i* dengan objek *j*

x_{ik} = nilai objek *i* pada peubah ke-*k*

x_{jk} = nilai objek *j* pada peubah ke-*k*

S = matriks kovarian

(Mongi, 2015)

2.3.4 Jarak *Log-Likelihood*

Jarak *log-likelihood* adalah jarak yang digunakan untuk peubah skala kontinu dan kategorik. Jarak antara *cluster j* dengan *cluster s* dapat didefinisiakan sebagai berikut:

$$d(j,s) = \xi_j + \xi_s + \xi_{(j,s)} \tag{2.5}$$

dengan:

$$\xi_j = -N\left(\sum_{k=1}^{K^A} \frac{1}{2}\log(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) - \sum_{k=1}^{K^B}\sum_{l=1}^{L_k} \frac{N_{jkl}}{N_j}\log\left(\frac{N_{jkl}}{N_j}\right)\right)$$

$$\xi_s = -N\left(\sum_{k=1}^{K^A} \frac{1}{2}\log(\hat{\sigma}_k^2 + \hat{\sigma}_{sk}^2) - \sum_{k=1}^{K^B}\sum_{l=1}^{L_k} \frac{N_{skl}}{N_j}\log\left(\frac{N_{skl}}{N_j}\right)\right)$$

$$\xi_{(js)} = -N\left(\sum_{k=1}^{K^A} \frac{1}{2}\log(\hat{\sigma}_k^2 + \hat{\sigma}_{(js)k}^2) - \sum_{k=1}^{K^B}\sum_{l=1}^{L_k} \frac{N_{(js)kl}}{N_j}\log\left(\frac{N_{(js)kl}}{N_j}\right)\right)$$

dengan:

N = banyaknya objek

N_j = jumlah objek di dalam *cluster j*

N_{jkl} = jumlah objek di *cluster j* untuk peubah kategorik ke-*k* dengan kategori ke-*l*

$\hat{\sigma}_k^2$ = ragam dugaan untuk peubah kontinu ke-*k* untuk keseluruhan objek

$\hat{\sigma}_{jk}^2$ = ragam dugaan untuk peubah kontinu ke-*k* untuk keseluruhan objek dalam *cluster j*

K^A = banyaknya peubah kontinu

K^B = banyaknya peubah kategorik

L_k = banyaknya kategori untuk peubah kategorik ke-*k*

(Mongi, 2015)

2.3.5 Jarak *Chi-Square*

Menurut Greenacre and Primicerio (2013), jarak *chi-square* adalah jarak untuk menghitung antara profil dalam jarak *Euclidean* terboboti menggunakan invers proporsi rata-rata sebagai bobot. Misal c_j menunjukkan elemen ke-*j* dari rata-rata profil yang merupakan kelimpahan proporsi peubah *j* dari seluruh kumpulan data. Maka jarak *chi-square* dinotasikan χ, antara dua objek dengan profil $x = [x_1 x_2 \dots x_j]$dan $y = [y_1 y_2 \dots y_j]$ dapat didefiniskan sebagai berikut:

$$\chi_{x,y} = \sqrt{\sum_{j=1}^{J} \frac{1}{c_j} (x_j - y_j)^2} \qquad (2.6)$$

2.4 Metode Pengelompokkan

2.4.1 Metode Hierarki

Metode hierarki *(hierarchial method)* adalah metode yang digunakan jika banyaknya *cluster* yang akan dibentuk belum ada informasi sebelumnya. Metode hierarki dilakukan dengan cara mengelompokkan dua atau lebih objek yang memiliki kesamaan terdekat, proses tersebut terus berlanjut hingga ke objek-objek lain yang memiliki kesamaan dan hasil akhir *cluster* yang terbentuk yakni berupa diagram pohon yang memiliki tingkatan antar objek berdasarkan tingkat kemiripan pada objek tersebut (Setiawan dan Pratiwi, 2019).

Metode hierarki dibagi menjadi dua yaitu metode penggabungan (*agglomerative method*) dan metode pemisahan (*divisive method*). Metode penggabungan yakni di awali dengan *n* buah *cluster* yang masing – masing anggota memiliki satu objek. Selanjutnya dua *cluster* yang terdekat digabungkan dan ditentukan kembali kedekatan antar *cluster* yang baru. Proses ini dilanjutkan sampai diperoleh satu *cluster* yang beranggota seluruh objek. Metode pemisahan adalah metode yang diawali denga satu *cluster* yang beranggota seluruh objek, selanjutnya objek-objek terjauh dipisahkan dan membentuk *cluster* lain. Proses ini terus berlanjut hinga seluruh objek masing-masing akan membentuk satu *cluster.* Jenis peubah yang digunakan dalam pengklasteran di metode hierarki adalah peubah kontinu (rasio dan interval) dan pengukutan jarak yang sering dipakai dalam metode hierarki yakni jarak *euclidean* atau jarak mahalanobis.

Menurut Rencher (2002), metode ini dibagi menjadi enam metode yaitu:

1. *Single Linkage*

   Single linkage disebut juga metode tetangga terdekat, jarak antara dua *cluster* A dan B didefinisikan sebagai jarak minimum antara titik A dan titik B:

   $$d(A,B) = min\{d(y_i, y_j)\ untuk\ y_i \in A\ dan\ y_j \in B\} \quad (2.7)$$

2. *Complete Linkage*

   Complete linkage disebut juga metode tetangga terjauh, jarak antara dua *cluster* A dan B didefinisikan sebagai jarak maksimum antara titik A dan titik B:

   $$d(A,B) = max\{d(y_i, y_j)\ untuk\ y_i \in A\ dan\ y_j \in B\} \quad (2.8)$$

3. *Average Linkage*

   Jarak antara dua *cluster* A dan B didefinisikan sebagai rata-rata jarak $n_A n_B$ antara titik n_A di A dan titik n_B di B:

   $$d(A,B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j-1}^{n_B} d(y_i, y_j) \quad (2.9)$$

   dimana jumlahnya adalah semua y_i di A dan semua y_j di B.

4. *Centroid*

   Pada metode *centroid*, jarak antara dua cluster A dan B di definisikan jarak *euclidean* antara vektor mean (sering disebut *centroid*) dari dua *cluster*:

   $$d(A,B) = d(\bar{y}_A, \bar{y}_B) \quad (3.0)$$

dimana,

$$\bar{y}_A = \sum_{i=1}^{n_A} \frac{y_i}{n_A} \qquad \bar{y}_B = \sum_{j=1}^{n_B} \frac{y_j}{n_B}$$

sehingga,

$$\bar{y}_{AB} = \frac{n_A \bar{y}_A + n_B \bar{y}_B}{n_A + n_B}$$

5. Median

   Jika dua *cluster* A dan B digabung menggunakan metode *centroid,* dan jika A berisi lebih banyak item daripada B, maka *centroid* baru $\bar{y}_{AB} = (n_A \bar{y}_A + n_B \bar{y}_B)/(n_A + n_B)$ kemungkinan lebih dekat ke $\bar{y}_A$ daripada $\bar{y}_B$. Untuk menghindari pembobotan vektor mean berdasarkan ukuran *cluster,* kita bisa menggunakan median garis mengikuti A dan B sebagai titik untuk menghitung *cluster* baru ke *cluster* lain:

$$m_{AB} = \frac{1}{2}(\bar{y}_A + \bar{y}_B) \qquad (3.1)$$

6. Metode Ward

   Ward (1963*)* mengusulkan penggunaan metode yang didasarkan pada hasil informasi yang minimum dari kenaikan pada jumlah kuadrat deviasi rataan *cluster*. Proses berhenti pada kenaikan yang menyebabkan *error sum of squares (ESS)* dari gabungan tiap *cluster* yang mungkin. Nilai ESS digunakan sebagai fungsi obyektif dan didefinisikan sebagai berikut:

$$ESS = \sum_{j=1}^{k} \left(\sum_{i=1}^{n_j} {x_{ij}}^2 - \frac{1}{n_j} \left(\sum_{i=1}^{n_j} x_{ij} \right)^2 \right) \qquad (3.2)$$

   Metode ini juga dikenal dengan metode varian minimum dan harus menggunakan jarak *Euclidean* namun sulit untuk menggunakannya tanpa bantuan komputer.

2.4.2 Metode Non-Hierarki

Metode non-hierarki adalah metode yang digunakan jika banyaknya *cluster* yang akan dibentuk sudah diketahui informasi sebelumnya. Metode non-hierarki di katakan sebagai metode *k-means.* Proses metode non-hierarki dilakukan dengan memilih sejumlah *k*, yaitu banyaknya *cluster*. Pemilihan *k* dilakukan secara

subyektif berdasarkan pengamatan bidang masing-masing. Jenis peubah yang digunakan dalam pengklasteran di metode ini adalah peubah kontinu (rasio dan interval) dan pengukutan jarak yang sering digunakan dalam metode non hierarki yakni jarak *Euclidean*. Menurut Lathifaturrahmah (2014), dalam penelitiannya data yang digunakan untuk dianalisis dengan metode non-hierarki yakni data/objek dengan ukuran sampel besar.

Menurut Ediyanto, dkk (2013), langkah-langkah algoritma *k-means cluster analysis* yaitu sebagai berikut:

1. Tentukan jumlah *cluster.*
2. Alokasikan objek ke dalam *cluster* secara random.
3. Hitung *centroid* sampel yang ada di masing-masing *cluster.*
4. Alokasikan masing-masing objek ke *centroid* ke terdekat.
5. Kembali ke langkah 3 apabila masih ada objek yang berpindah *cluster* atau masih ada perubahan nilai *centroid,* ada yang di atas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan.

2.4.3 Metode *Two Step Cluster*

Metode *two step cluster* adalah metode yang dirancang untuk mengatasi jumlah objek dengan ukuran yang lebih besar, terutama pada masalah objek dengan skala pengukuran yang berbeda yaitu peubah kontinu dan kategorik. Menurut Li and Sun (2011), analisis *two-step cluster* merupakan salah satu metode *cluster* yang dirancang untuk menyingkapkan *cluster* alami dari kumpulan data yang sebelumnya tidak terlihat. *Two-step cluster* memberikan informasi tentang pentingnya setiap peubah dalam pembangunan *cluster* tertentu, dimana hal tersebut merupakan fitur menarik dibandingkan dengan metode pengklasteran tradisional.

Ukuran jarak yang digunakan pada metode *two step cluster* adalah jarak *log-likehood* untuk skala data campuran yaitu numerik dan kategorik atau jarak *Euclidean* hanya untuk skala data berupa numerik. Ada 3 asumsi yang mendasari ukuran jarak *log-likehood* yaitu paubah saling bebas, peubah kategorik diasumsikan berdistribusi multinomial dan peubah numerik diasumsikan berdistribsi normal (Setiawan dan Pratiwi, 2019).

Menurut Bacher, *et al* (2004), Proses pengklasteran pada metode *two step cluster* memiliki dua tahap yaitu pembentukan *cluster* awal *(pre-clustering)* dan pembentukan *cluster* optimal.

1. Pembentukan *Cluster* Awal *(Pre-Clustering)*

Langkah pertama di metode *two step cluster* ialah pembentukan *cluster* awal *(pre-clustering)* dilakukan dengan pendekatan secara sekuensial, yaitu setiap objek yang diamati satu persatu bersumber pada ukuran jarak dan setelah itu ditetapkan apakah objek tersebut akan bergabung dalam *cluster* yang sudah terbentuk atau membentuk *cluster* baru.

Menurut Schiopu (2010), pada pendekatan ini diimplementasikan dengan membentuk *Cluster Feature (CF) Tree. CF Tree* terdiri dari tingkatan cabang *(node)* serta tiap-tiap cabang berisikan beberapa objek/data yang dientrikan. Jika dimisalkan dengan sebuah pohon, sehingga tingkatan cabang tersebut terdiri dari batang pohon, dahan serta daun. Pada *CF Tree* terdapat tingkatan daun yang disebut sebagai daun entri yang berada pada cabang merepresentasikan hasil *sub-cluster* atau anak *cluster*.

Menurut Mongi (2015), prosedur *CF Tree* dapat dilakukan dengan cara memilih satu amatan awal secara random untuk diukur jaraknya masing-masing amatan dengan amatan lain berdasarkan ukuran jarak yang sudah ditetapkan. Apabila

besaran jarak letaknya di dalam daerah penerimaan, selanjutnya amatan tersebut akan masuk kedalam anggota anak *cluster*. Apabila besaran jarak letaknya di luar daerah penerimaan, maka amatan tersebut akan masuk kedalam *cluster* yang telah terbentuk atau yang nantinya akan menjadi daun entri baru.

Apabila dalam suatu kasus data/objek ditemui pencilan, maka harus diperiksa apakah *CF Tree* yang terbentuk dapat dimasukkan kedalam *cluster* yang telah terbentuk tanpa harus membentuk lagi *CF Tree* baru. Untuk mendeteksi pencilan dapat dilakukan dengan perhitungan pada jarak *log-likelihood*. Apabila terdapat jarak terbesar antar *cluster* yang melebihi nilai titik kritis *C*, yaitu:

$$C = \log(V) \tag{3.3}$$

$$V = \prod_k R_k \prod_m L_m \tag{3.4}$$

dengan:

R_k = range dari peubah kontinu ke-*k*

L_m = banyaknya kategorik untuk peubah kategorik ke-*m*

Sedangkan pada jarak *Euclidean,* dikatakan data yang memuat pencilan jika jarak terbesar antar *cluster* melebihi nilai di titik kritis C, rumus C dapat didefinisikan:

$$C = 2\left(\sum_{i=1}^{K^A} \frac{\hat{\sigma}_{kl}^2}{K^A}\right)^{\frac{1}{2}} \tag{3.5}$$

dengan:

K^A = banyaknya peubah kontinu

$\hat{\sigma}_{kl}^2$ = ragam dugaan untuk peubah kontinu ke-*l* dalam *cluster* ke-*k*

Pembentukan *CF Tree* terdiri dari dua tahap. Tahapan yang pertama adalah tahap penyisipan *(inserting)* dan tahapan kedua yaitu tahap pembentukan kembali *(rebuilding)*. Pada tahapan penyisipan dilakukan secara acak dipilih satu objek kemudian diukur jaraknya ke objek yang lain. Apabila jarak tersebut hasil nilainya kurang dari jarak maksimum, maka objek tersebut akan dimaksudkan kedalam

satu *cluster*. Sedangkan apabila jarak tersebut hasil nilainya melebihi jarak maksimum, maka objek tersebut dikatakan pencilan.

Pada tahapan pembentukan kembali dilakukan berdasarkan objek yang dianggap sebagai pencilan, selanjutnya akan dibentuk suatu *cluster* baru. Apabila *CF Tree* bertambah banyak sehingga melewati batas ukuran maksimum yang telah ditentukan, sehingga batas jarak maksimum perlu ditingkatkan agar dapat memasukkan banyak objek. Peningkatan jarak ini dapat mengakibatkan objek-objek yang berasal dari *cluster* yang berbeda bergabung menjadi satu *CF,* selanjutnya dihasilkan *CF Tree* dengan ukuran lebih kecil dari sebelumnya. Hasil yang diperoleh dari *CF Tree* diklasterkan dengan analisis *cluster* berhierarki dengan metode penggabungan. Tiap-tiap *cluster* yang terbentuk di tahap pertama akan digabungkan satu persatu berdasarkan ukuran yang telah ditentukan. Prosedur ini akan berakhir sampai seluruh *sub-cluster* mejadi satu *cluster*.

2. Pembentukan *Cluster* Optimal

Menurut Mongi (2015), tahap selanjutnya ialah melakukan tahapan pembentukan *cluster* optimal, dalam menentukan banyaknya jumlah *cluster* dapat dilakukan dengan cara dua tahap, tahap pertama adalah menghitung *Bayesian Information Criterion* (BIC) atau *Akaike Information Criterion* (AIC) untuk setiap *cluster*. Rumus BIC dan AIC untuk *cluster J* dapat didefinisikan sebagai berikut:

$$BIC(J) = -2\sum_{j=1}^{J}\xi_j + m_j \log(N) \tag{3.6}$$

$$AIC(J) = -2\sum_{j=1}^{J}\xi_j + 2m_j \tag{3.7}$$

$$m_j = J\left\{2K^A + \sum_{k=1}^{K^B}(L_k - 1)\right\}$$

$$\xi_j = -N\left(\sum_{k=1}^{K^A}\frac{1}{2}\log\left(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2\right)\sum_{k=1}^{K^B}\sum_{k=1}^{L_k}\frac{N_{jkl}}{N_j}\log\left(\frac{N_{jkl}}{N_j}\right)\right)$$

dengan:

$N =$ jumlah data

$N_j =$ jumlah data di dalam *cluster j*

N_{jkl} = jumlah data di *cluster j* untuk peubah kategorik ke-*k* dengan ke-*l*

$\hat{\sigma}_k^2$ = ragam dugaan untuk peubah kontinu ke-*k* untuk keseluruhan data

$\hat{\sigma}_{jk}^2$ = ragam dugaan untuk peubah kontinu ke-*k* dalam *cluster j*

K^A = jumlah peubah kontinu

K^B = jumlah peubah kategorik

L_k = jumlah kategorik untuk peubah kategorik ke-*k*

Selanjutnya hasil pada perhitungan tersebut akan digunakan untuk menduga jumlah *cluster*. Tahap kedua adalah mencari jarak terbesar antara dua *cluster* yang saling berdeketan pada tiap-tiap tahapan dalam pengklasteran. Solusi banyaknya *cluster* terbaik adalah BIC yang memiliki nilai terkecil, namun ada beberapa masalah di dalam pengklasteran dimana nilai BIC akan terus bertambah jika jumlah *cluster* semakin meningkat. Sehingga dalam kasus tersebut, rasio perubahan BIC (*ratio BIC changes*) dan rasio perubahan jarak (*ratio of distance measure changes*) yang mana nilai tersebut untuk mengidentifikasi solusi dari banyaknya *cluster* optimal. Solusi untuk banyaknya *cluster* optimal akan memiliki rasio perubahan BIC dan rasio perubahan jarak yang besar.

Perbandingan antar jarak untuk *k cluster* digunakan untuk mengetahui banyaknya *cluster* yang terbentuk, dengan rumus perbandingan sebagai berikut :

$$R(k) = \frac{d_{k-1}}{d_k} \tag{3.8}$$

$$d_k = l_{k-1} - l_k$$

$$l_v = \frac{(r_v \log n - BIC_v)}{2} \tag{3.9}$$

$$l_v = \frac{(2r_v - AIC_v)}{2}$$

$$v = k, k-1$$

dengan:

d_{k-1} = jarak jika *k cluster* digabungkan dengan $k-1$ *cluster*

$R(k)$ = rasio perubahan jarak

(Setiawan dan Pratiwi, 2019)

2.5 Peubah yang Berpengaruh

Menurut Schiopu (2010), untuk menentukan peubah yang berpengaruh dapat dilakukan dengan cara uji *t-student* untuk peubah numerik dan uji *chi-square* untuk peubah kategorik.

2.5.1 Uji *T-Student*

Pada uji *t-student* yang digunakan adalah rataan dari karakteristik di dalam *cluster* terhadap rataan umum (Bustami, dkk., 2014). Uji *t-student* mengevaluasi perbedaan antara rata-rata dua kelompok independen atau tidak terikat. Uji mengevaluasi apakah rata-rata untuk dua kelompok independen berbeda secara signifikan satu sama lain (Mohamed and Awang, 2015).

Hipotesisi uji *t-student* yaitu :

$H_0: \mu_{jk} = \mu_k$ (peubah numerik tidak berpengaruh terhadap *cluster*)

$H_1: \mu_{jk} \neq \mu_k$ (peubah numerik berpengaruh terhadap *cluster*)

Dengan tingkat signifikansi α adalah 0,05 % dan statistik uji *t* hitung yaitu:

$$t_{hit} = \frac{\hat{\mu}_k - \hat{\mu}_{jk}}{\hat{\sigma}_{jk}/\sqrt{N_k}} \qquad (4.0)$$

dengan:

$\hat{\mu}_k$ = rataan peubah numerik ke-k

$\hat{\mu}_{jk}$ = rataan peubah numerik ke-k kelompok ke-j

$\hat{\sigma}_{jk}$ = simpangan baku peubah numerik ke-k pada kelompok ke-j

N_k = jumlah observasi pada peubah numerik ke-k

Jika nilai $t_{hit} > t_{tabel}$ maka hipotesis nol ditolak, artinya peubah numerik tersebut berpengaruh terhadap *cluster*.

2.5.2 Uji *Chi-Square*

Pada uji *chi-square* yang digunakan adalah proporsi individu atau objek antar kategorik dalam satu *cluster* untuk melihat ada tidaknya hubungan antar peubah pada *cluster* (Sumertajaya, dkk., 2007). Uji *chi-square* adalah uji statistik yang digunakan untuk menguji independensi dan kecocokan (Mohamed and Awang, 2015).

Hipotesis uji *chi-square* yaitu:
$H_0: \pi_{jkl} = \pi_{kl}$ (peubah kategorik tidak berpengaruh terhadap *cluster*)
$H_1: \pi_{jkl} \neq \pi_{kl}$ (peubah kategorik berpengaruh terhadap *cluster*)
Dengan tingkat signifikansi α adalah 0,05 % dan statistik uji *chi-square* hitung yaitu:

$$\chi^2 = \sum_{l=1}^{l_k} \left(\frac{N_{jkl} - N_{kl}}{N_{kl}}\right)^2 \quad (4.1)$$

dengan:
N_{jkl} = jumlah observasi pada *cluster* ke-j untuk peubah kategorik ke-k dengan kategorik ke-l
N_{kl} = jumlah observasi pada *cluster* untuk peubah kategorik ke-k dengan kategorik ke-l
l_k = jumlah kategorik untuk peubah kategorik ke-k
Jika nilai $\chi^2{}_{hit} > \chi^2{}_{tabel}$ maka hipotesis nol ditolak, artinya peubah kategorik tersebut berpengaruh terhadap *cluster*.

2.6 Validitas *Clustering*

Untuk melihat validitas *clustering*, perlu dilakukan pengujian dengan *silhouette*. Koefisien *silhouette* dapat digunakan untuk memeriksa objek di dalam setiap *cluster*. Koefisien ini nilainya berkisar dari -1 sampai 1. Nilai negatif dari koefesien *silhouette* berarti bahwa objek yang diukur lebih berhubungan dengan

objek pada *cluster* lainnya. Semakin rendah nilainya dari koefesien *silhouette* semakin jauh jarak antara objek dan *cluster* (Jauhiainen and Tommi, 2017).

Menurut Kauffman and Roesseeuw (1990), kriteria subjektif pengukuran pengelompokkan berdasarkan nilai *Silhoutte Coefficient (SC),* dapat dilihat pada tabel berikut untuk rentang nilai dan kriteria.

Tabel 1. Kriteria Pengukuran Koefisien *Silhoutte*

Nilai SC	Kriteria
0.71 – 1.0	Struktur kuat
0.51 – 0.70	Struktur baik
0.26 – 0.50	Struktur lemah
< 0.25	Struktur buruk

Menurut Supandi, *et al* (2021), untuk rata-rata koefisien *silhouette* dapat digunakan untuk mengukur kualitas klasterisasi. Nilainya juga berkisar antara -1 hingga 1. Jika nilai koefisien *silhouette* antara -1 hingga 0,2 maka kualitasnya diklasifikasikan buruk, 0,2 sampai 0,5 berada pada kondisi cukup, dan 0,5 sampai 1 berada pada kondisi baik. Rumus koefisien *silhouette* dapat dilihat sebagai berikut:

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \tag{4.2}$$

$$\bar{S} = \frac{1}{N}\sum_{i=1}^{N} S_i \tag{4.3}$$

dengan:

S_i = koefisien *silhouette* untuk objek ke-i

b_i = rata-rata jarak minimum antara objek ke-i dalam *cluster* yang berbeda

a_i = rata-rata jarak minimum antara objek ke-i dalam *cluster* yang sama

$\bar{S}$ = nilai rata-rata untuk koefisien *silhouette*

N = jumlah total pengamatan

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester ganjil tahun 2021/2022 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder berupa data sosial dan kependudukan di kabupaten/kota Provinsi Jawa Barat pada tahun 2019 yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Barat (Lampiran 1).

Adapun jumlah kabupaten/kota Provinsi Jawa Barat yaitu sebanyak 27 kabupaten/kota terdiri dari Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majelengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kabupaten Pengandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya dan Kota Banjar.

Pengumpulan data berdasarkan peubah yang digunakan dalam penelitian ini, berupa peubah campuran yaitu kategorik dan numerik. Adapun peubah-peubah yang digunakan adalah sebagai berikut.

Peubah kategorik terdiri dari:

a. Status kabupaten (X_1) yaitu kota madya dan kabupaten
b. Status bencana (X_2) yaitu kategori rawan dan kategori sedang

Peubah numerik terdiri dari:

a. Jumlah penduduk miskin (X_3)
b. Jumlah fasilitas kesehatan (X_4)
c. Jumlah rumah tangga (X_5)
d. Jumlah sekolah dasar (X_6)

3.3 Metode Penelitian

Penelitian ini dilaksanakan dengan studi pustaka yaitu dengan pengkajian secara teoritis dan praktik komputasi. Adapun langkah-langkahnya yang harus dilalui untuk melakukan penelitian dengan metode *two step cluster* adalah sebagai berikut:

1. Menentukan peubah dan objek yang akan digunakan dalam penelitian.
   Adapun peubah yang digunakan yaitu peubah numerik dan kategorik.

2. Melakukan stardarisasi data
   Standarisasi pada data dilaksanakan agar peubah yang digunakan pada penelitian ini memiliki satuan pengukuran yang sama.

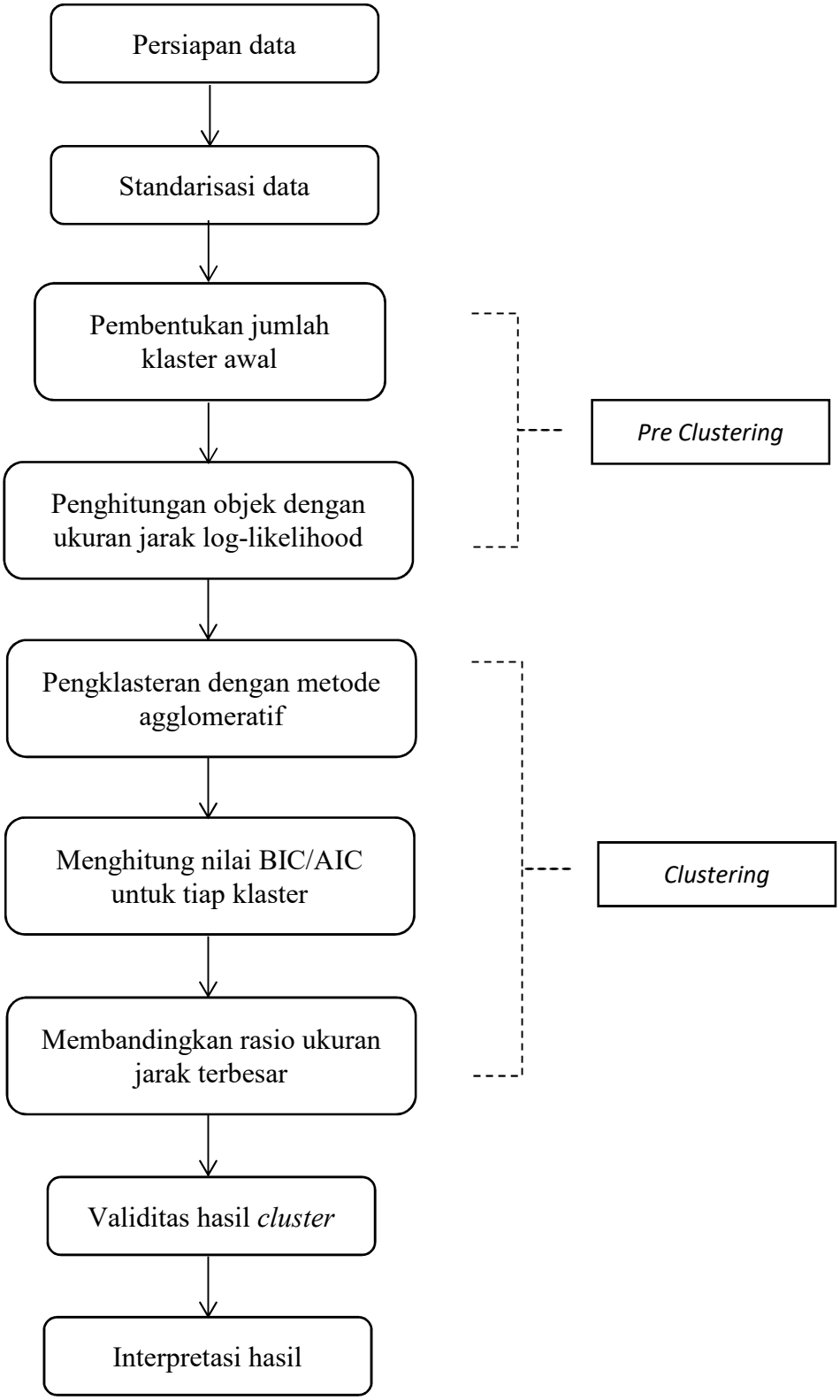
3. Melakukan prosedur analisis *cluster*
   Pada penelitian ini digunakan pengklasteran dengan metode *two step cluster* terhadap peubah-peubah kategorik dan kontinu, dengan langkah-langkah sebagai berikut:

a) Tahap pertama
    1. Dilakukan pemilihan objek agar dapat dijadikan *sub-cluster* awal secara acak, selanjutnya menentukan jumlah batas maksimum objek agar dapat begabung dengan *sub-cluster*.
    2. Memasukkan objek secara satu persatu agar terbentuk *sub-cluster* dengan cara menghitung jarak berdasarkan jarak *log-likehood. Sub-cluster* terdiri dari campuran objek anggota-anggota yang masuk pada bagian *sub-cluster*.
    3. Penggabungan terus dilakukan hingga semua objek anggota bergabung dalam suatu *sub-cluster* sehingga diperoleh pembentukan *cluster* awal yang mana merupakan hasil dari pembentukan sub-sub *cluster* pada tahap pertama.

b) Tahap kedua
    1. Dilakukan penghitungan nilai tengah sub-sub *cluster* yang terbentuk di tahap pertama.
    2. Hasil dari nilai tengah sub-sub *cluster* di tahap pertama kemudian diklasterkan dengan menggunakan metode hierarki terkhusus metode penggabungan.
    3. Selanjutnya dilakukan pemilihan jumlah *cluster* optimum yang dihasilkan dari pembentukan *Cluster Feature Tree*. Pemilihan *cluster* dilakukan berdasarkan pada kriteria *BIC* terkecil atau dengan melihat rasio perubahan jarak terbesar .
    4. Dihasilkan *cluster-cluster* yang terbentuk.

4. Melakukan interpretasi karakteristik *cluster-cluster* yang terbentuk
   Setelah *cluster* terbentuk maka tahap selanjutnya ialah memberi ciri spesifik untuk menggambarkan isi *cluster* tersebut.

5. Melakukan pemeriksaan kekuatan pembagian *cluster* atau mengukur kualitas klasterisasi dengan menggunakan koefisien *silhouette.*



Gambar 1. Diagram Alir Metodologi Penelitian

V. KESIMPULAN

Berdasarkan penelitian diatas dapat disimpulkan yaitu sebagai berikut:

1. Hasil pengklasteran *two step cluster* pada data sosial dan kependudukan provinsi Jawa Barat tahun 2019 dengan menggunakan 6 peubah yang terdiri dari 2 peubah kategorik dan 4 peubah numerik menghasilkan 2 *cluster* optimal.
2. *Cluster* satu dicirikan memiliki karakteristik dengan jumlah anggota terbanyak yaitu 18 berstatus kabupaten. Peubah yang berpengaruh terhadap pembentukan *cluster* pertama adalah status kabupaten dan pembentukan *cluster* pertama tidak berdasarkan peubah-peubah kontinu yang digunakan.
3. *Cluster* dua dicirikan dengan karakteristik dengan jumlah anggota sedikit yaitu 9 berstatus kota madya. Peubah kategorik yang mempengaruhi pembentukan *cluster* dua yaitu status kabupaten dan status bencana. Adapun peubah kontinu yang mempengaruhi *cluster* dua yaitu jumlah penduduk miskin, jumlah sekolah dasar dan jumlah fasilitas kesehatan.
4. *Cluster* 1 memiliki keadaan sosial lebih baik daripada *cluster* 2, sehingga pemerintah harus meningkatkan keadaan sosial pada *cluster* 2 agar kabupaten/kota yang berada di *cluster* 2 dapat meningkatkan seperti sarana, prasarana, sosial dan ekonomi untuk mendapatkan kesejahteraan secara merata di Provinsi Jawa Barat.

DAFTAR PUSTAKA

Bacher, J. Wenzig, K. and Vogler, M. 2004. *SPSS Two Step Cluster - A First Evaluation*. Arbeits- und Diskussionspapierre.

Bustami. Abdullah, D. dan Fadlisyah. 2014. *Statistika Terapannya Pada Bidang Informatika Edisi Pertama.* Graha Ilmu, Yogjakarta.

Ediyanto, Mara, M. N. dan Satyahadewi, N. 2013. Pengklasifikasian Karakteristik dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)*. **2**(2):133-136.

Hair, J. F. Jr. Black, W. C. Babin, B. J. And Anderson, R. E. 2019. *Multivariate Data Analysis Eighth Edition*. Cengage Learning EMEA, United Kingdom.

Hapsari, I. A. Kusnandar, D. dan Imro'ah. N. 2020. Metode Two Step Cluster Dalam Mengelompokkan Mahasiswa Fmipa Untan. *Jurnal Ilmiah Math. Stat. dan Terapannya.* **9**(1):73-180.

Li, H. and Sun, J. 2011. Mining Business Failure Predictive Knowledge Using Two-Step Clustering. *African Journal of Business Management.* **5(**11):4107-4120 .

Jauhiainen, S. dan Tommi, K. 2017. A Simple Cluster Validation Index with Maximal Coverage. *ESANN 2017 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*:293-298.

Kauffman, L and Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis* , USA: Wiley Series in Probability and Statistics, 1990.

Lathifaturrahmah. 2014. Perbandingan Hasil Penggerombolan K-Means, Fuzzy K- Means, dan Two Step Clustering. *JPM IAIN Antasari.* **2**(1):39-62.

Greenacre, M. and Primicerio, R. 2013. *Multivariate Analysis of Ecological Data*. Fundacion BBVA, Plaza de San Nicolas, 4, 48005, Bilbao.

Mohamed, N. and Awang, S. R. 2015. The Multiple Intelligence Classification of Management Graduates Using Two-Step Cluster Analysis. *Malaysian Journal of Fundamental and Applied Sciences*. **11**(1):48-51.

Mongi C. E. 2015. Penggunaan Analisis Two Step Clustering Untuk Data Campuran. *Jurnal de Cartesian (JdC)*. **4**(1):9-19.

Nugroho, S. 2008. *Statistika Multivariate Terapan*. UNIB Press, Bengkulu.

Rachmatin, D. 2014. Aplikasi Metode-Metode Agglomerative Dalam Analisis Cluster Pada Data Tingkat Polusi Udara. *Jurnal Ilmiah Program Studi Matematika STKIP Siliwangi Bandung*. **2**(3):133-149.

Rencher, A. C. 2002. *Method of Multivariate Analysis*. John Wiley & Sons, Inc. Canada.

Setiawan, A. H. dan Pratiwi, N. 2019. Penerapam Metode Two Step Cluster Untuk Pengelompokkan Potensi Desa. *Jurnal Statistika Industri dan Komputasi.* **4**(2):41-51.

Schiopu, D. 2010. *Applying two step cluster analysis for identifying bank customers' profile*. *EI-TC*. **62**(3):66-75.

Sumertajaya I. M. Erfiani dan Putri W. D. Y. 2007. Analisis Gerombol Menggunakan Metode Two Step Cluster. *Forum Statistika dan Komputasi*. **12**(1):18-23.

Supandi, A. Saefuddin, A. and Sulvianti, I. D. 2021. Two Step Cluster Application to Classify Villages in Kabupaten Madiun Based on Village Potential Data. *Journal of Statistics.* **10**(1):12-26.