

**PERBANDINGAN ANALISIS KLASTER *K-MEANS* DAN
AVERAGE LINKAGE PADA PENGELOMPOKKAN
KEMISKINAN DI PROVINSI JAWA BARAT PADA TAHUN**

2022

SKRIPSI



Diajukan oleh:

NABILA GARISOH ALFARISI

NIM: 1808046010

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI WALISONGO SEMARANG**

2023

PERNYATAAN KEASLIAN

Yang bertandatangan di bawah ini:

Nama : Nabila Garisoh Alfalisi

NIM : 1808046010

Jurusan : Matematika

Menyatakan bahwa skripsi ini yang berjudul:

**Perbandingan Analisis Kluster *K-means* dan *Average Linkage*
Pada Pengelompokan Kemiskinan Di Provinsi Jawa Barat Pada
Tahun 2022**

Secara keseluruhan adalah hasil penelitian/karya sendiri, kecuali bagian tertentu yang dirujuk sumbernya.

Semarang, 11 Oktober 2023

Pembuat Pernyataan



Nabila Garisoh Alfalisi

NIM. 1808046010

HALAMAN PENGESAHAN



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI WALISONGO
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Prof. Dr. Hamka Kampus II Ngaliyan Semarang 50185 Telp. 7601295 Fax. 7615387

PENGESAHAN

Naskah skripsi berikut ini :

Nama : Nabila Garisoh Alfalisi

NIM : 1808046010

Jurusan : Matematika

Fakultas : Sains dan Teknologi

Judul : **Perbandingan Analisis Kluster *K-means* dan *Average Linkage* Pada Pengelompokkan Kemiskinan Di Provinsi Jawa Barat**

Telah di ujian dalam sidang *munaqosyah* oleh Dewan Penguji Fakultas Sains dan Teknologi Universitas Islam Negeri Walisongo Semarang dan dinyatakan LULUS pada tanggal : **14 Desember 2022.**

Dan dapat diterima sebagai syarat guna memperoleh gelar Sarjana Strata 1 tahun akademik 2022/2023

Semarang, 14 Desember 2022

DEWAN PENGUJI

Ketua Sidang,

Ayus Kiana Isnawati, M.Sc
NIP. 198510192019032014

Penguji Utama I,

Siti Maslilah, M.Si
NIP. 197706112011012004

Pembimbing I,

Eva Khoirun Nisa, S.Si., M.Si.
NIP. 198701022019032010

Sekretaris Sidang,

Eva Khoirun Nisa, S.Si., M.Si.
NIP. 198701022019032010

Penguji II,

Seftina Diah Miasary, M.Sc.
NIP. 198709212019032010

Pembimbing II,

Ariska Kurnia Rachmawati, M.Sc
NIP. 198908112019032019

NOTA PEMBIMBING I

Semarang, 11 Oktober 2023

Yth. Ketua Program Studi Matematika Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum wr.wb.

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan,
arahan, dan koreksi naskah skripsi dengan:

Judul : Perbandingan Analisis Kluster *K-means* dan *Average
Linkage* Pada Pengelompokkan Kemiskinan DI
Provinsi Jawa Barat
Nama : Nabila Garisoh Alfarisi
NIM : 1808046010
Program Studi : Matematika

Saya memandang bahwa naskah skripsi tersebut sudah dapat
diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo
Semarang untuk diajukan dalam Sidang Munaqasyah.

Wassalamu'alaikum wr.wb.

Pembimbing I,



Eva Khoirun Nisa, S.Si., M.Si.
NIP. 198701022019032010

NOTA PEMBIMBING II

Semarang, 11 Oktober 2023

Yth. Ketua Program Studi Matematika Fakultas Sains dan Teknologi
UIN Walisongo Semarang

Assalamu'alaikum wr.wb.

Dengan ini diberitahukan bahwa saya telah melakukan bimbingan,
arahan, dan koreksi naskah skripsi dengan:

Judul : Perbandingan Analisis Klaster *K-means* dan *Average Linkage* Pada Pengelompokkan Kemiskinan DI
Provinsi Jawa Barat

Nama : Nabila Garisoh Alfarisi

NIM : 1808046010

Program Studi : Matematika

Saya memandang bahwa naskah skripsi tersebut sudah dapat
diajukan kepada Fakultas Sains dan Teknologi UIN Walisongo
Semarang untuk diajukan dalam Sidang Munaqasyah.
Wassalamu'alaikum wr.wb.

Pembimbing II,



Ariska Kurnia Rachmawati, M.Sc

NIP. 19890811 201903 2 019

ABSTRAK

Kemiskinan adalah keadaan saat ketidakmampuan untuk memenuhi kebutuhan dasar seperti makanan, pakauan, tempat berlindung, pendidikan, dan Kesehatan. Tujuan dari penelitian ini adalah untuk mengelompokkan 27 kabupaten/kota di Jawa Barat dan mengetahui gambaran karakteristik setiap klaster yang terbentuk berdasarkan data kemiskinan di Jawa Barat pada tahun 2022. Metode yang digunakan dalam pengelompokkan adalah *K-means* dan *average linkage* dengan menggunakan 4 variabel numerik. Penentuan metode terbaik dilakukan dengan menghitung *silhouette coefficient*. Adapun tahapan dalam penelitian ini yaitu: (1) mengumpulkan data, (2) menentukan variabel, (3) melakukan analisis deskriptif, (4) standarisasi data, (5) melakukan uji *representative* dan uji korelasi, (6) perhitungan jarak *euclidean*, (7) proses pengelompokkan metode *K-means* dan *average linkage*, (8) interpretasi hasil analisis, dan (9) menentukan metode terbaik. Hasil penelitian menunjukkan jumlah klaster optimal pada *K-means* sebanyak 3 klaster yang terdiri dari klaster 1 beranggotakan 8 kabupaten/kota dengan karakteristik pengeluaran perkapita dan garis kemiskinan tinggi, klaster 2 beranggotakan 14 kabupaten/kota dengan karakteristik kemiskinan sedang, dan klaster 3 beranggotakan 5 kabupaten/kota dengan karakteristik jumlah penduduk miskin dan jumlah penduduk miskin ekstrem tinggi. Untuk metode *average linkage* jumlah klaster optimal sebanyak 2 klaster yang terdiri

dari klaster 1 beranggotakan 1 kabupaten/kota dengan karakteristik jumlah penduduk miskin dan jumlah penduduk miskin ekstrem tinggi dan klaster 2 beranggotakan 26 kabupaten/kota dengan karakteristik kemiskinan sedang. Metode terbaik pada penelitian ini adalah *average linkage* dengan nilai *silhouette coefficient* terbesar yaitu 0,6.

Kata Kunci: Analisis Klaster, *K-means* klaster, *Average Linkage*, Kemiskinan.

KATA PENGANTAR

Assalamu 'alaikum warahmatullahi wabarakatuh.

Dengan membaca *Alhamdulillah* sebagai bentuk rasa Syukur kehadiran Allah SWT atas segala rahmat, hidayah, serta karunia-Nya sehingga penulis dapat menyelesaikan penulisan skripsi ini yang berjudul, “Pebandingan Analisis Klaster K-means dan Average Linkage Pada Pengelompokkan Kemiskinan Di Provinsi Jawa Barat Pada Tahun 2022”. Tidak lupa sholawat serta salam kepada baginda Nabi Muhammad SAW yang selalu kita nantikan syafaatnya di hari akhir. Skripsi ini ditulis untuk memenuhi salah satu syarat memperoleh gelar Sarjana Strata (S1) pada program studi Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Walisongo Semarang. Dalam penulisan skripsi ini penulis menghadapi beberapa kesulitan, sehingga penulisannya tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Dr. H. Ismail, M.Ag., selaku Dekan Fakultas Sains dan Teknologi UIN Walisongo Semarang .
2. Ibu Emy Siswanah, M.Sc, selaku Ketua Program Studi Matematika Fakultas Sains dan Teknologi UIN Walisongo Semarang sekaligus dosen wali.

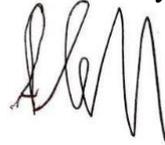
3. Bapak Aunur Rahman, M.Pd selaku Sekretaris Program Studi Matematika UIN Walisongo Semarang yang telah memberikan dukungan dalam penulisan skripsi ini.
4. Ibu Eva Khoirun Nisa, M.Si, selaku Dosen Pembimbing I dan Ibu Ariska Kurnia Rachmawati, M.Sc, selaku Dosen Pembimbing II yang telah meluangkan waktu dan pikiran serta memberikan bimbingan, pengarahan, dan saran yang bermanfaat dalam proses penulisan skripsi ini.
5. Bapak/Ibu Dosen dan Staf Program Studi Matematika UIN Walisongo Semarang yang telah mendidik, mengarahkan dan membantu proses skripsi ini.
6. Tim BPS Jawa Barat yang telah membantu dalam proses pengumpulan data skripsi ini.
7. Bapak Alim Priyanto, Ibu Ratmini, Mas Firda, Mas Ilham, Dek Bara dan keluarga tercinta yang senantiasa mendampingi Langkah penulis dengan doa restu dan memberikan perhatian, dukungan, serta semangat kepada penulis dalam menyelesaikan skripsi ini.
8. Teman-teman Angkatan 2018 Program Studi Matematika UIN Walisongo Semarang yang telah membantu dalam kegiatan selama kuliah dan telah memberikan keceriaan, semangat serta kebersamaannya selama ini.

9. Serta, kepada semua pihak yang telah membantu penulis dalam proses penulisan skripsi ini yang tidak dapat disebutkan Namanya satu per satu.

Penulis juga menyadari bahwa skripsi ini masih terdapat ketidaksempurnaan baik dari diksi, penulisan, dan penyajiannya. Oleh karena itu, penulis mengharapkan kritik dan saran dari pembaca untuk perbaikan skripsi ini. Akhir kata, semoga dengan adanya skripsi ini dapat bermanfaat bagi semua pihak.

Semarang, 11 Oktober 2023

Pembuat Pernyataan



Nabila Garisoh Alfarisi

NIM. 1808046010

DAFTAR ISI

PERNYATAAN KEASLIAN	ii
HALAMAN PENGESAHAN.....	iii
NOTA PEMBIMBING I	iv
NOTA PEMBIMBING II	v
ABSTRAK	vi
KATA PENGANTAR	viii
DAFTAR ISI.....	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	8
1.3 Tujuan Penelitian	9
1.4 Manfaat Penelitian	10
1.5 Batasan Masalah	11
BAB II LANDASAN PUSTAKA	12
2.1 Tinjauan Pustaka.....	12
2.1.1 Statistik Deskriptif.....	12
2.1.2 Analisis Multivariat.....	14
2.1.3 Data Mining.....	17
2.1.4 Standarisasi Data.....	20

2.1.5	Jarak <i>Euclidean</i>	21
2.1.6	Analisis Kluster.....	22
2.1.7	Algoritma <i>K-means</i>	26
2.1.8	<i>Silhouette Coefficient</i>	29
2.1.9	Metode <i>Elbow</i>	31
2.1.10	Metode <i>Gap Statistic</i>	33
2.1.11	Metode <i>Hierarki</i>	33
2.1.12	Metode <i>Average Linkage</i>	34
2.1.13	Pengertian Kemiskinan	35
2.1.14	Ciri Kemiskinan	37
2.2	Penelitian Terdahulu	39
2.3	Kerangka Berpikir	42
BAB III METODE PENELITIAN		46
3.1	Jenis Penelitian	46
3.2	Populasi dan Sampel Penelitian	46
3.3	Sumber Data	47
3.4	Variabel Penelitian	47
3.5	Teknik Analisis Data	49

BAB IV HASIL DAN PEMBAHASAN	52
4.1 Analisis Deskriptif	52
4.2 Standarisasi Data.....	55
4.3 Uji Asumsi Korelasi	57
4.4 Uji Asumsi <i>Representative</i>	61
4.5 Perhitungan Jarak	62
4.6 Analisis <i>K-means</i> Kluster	65
4.7 Metode <i>Average Linkage</i>	75
4.8 Perbandingan Hasil Kluster <i>K-means</i> dan <i>Average</i>	
<i>Linkage</i>	83
BAB V PENUTUP	85
5.1 Kesimpulan	85
5.2 Saran	88
DAFTAR PUSTAKA	89
LAMPIRAN-LAMPIRAN	102
RIWAYAT HIDUP.....	119

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemiskinan merupakan problem yang memiliki banyak aspek dan sangat *kompleks* (Widayati, 2013). Kemiskinan terjadi dikarenakan ketidakmampuan untuk mencukupi kebutuhan yang paling mendasar dan tidak mampu mendapatkan tingkat minimal dari tujuan yang ditargetkan (Nasmiwati & Triani, 2019). Tanda-tanda sosial kemiskinan yang muncul di berbagai negara antara lain pendapatan buruk, produktivitas rendah, dan keterbelakangan ilmu pengetahuan dan teknologi (Nasmiwati & Triani, 2019).

Karena masih banyak masyarakat miskin di Indonesia, kemiskinan masih menjadi permasalahan utama yaitu sebanyak 26,36 juta jiwa (Badan Pusat Statistik Indonesia, 2023). Menurut Badan Pusat Statistik Indonesia, (2023), 12,36% individu hidup dalam kemiskinan di pedesaan dan 7,53% di perkotaan, sebanyak 49,41% rumah tangga miskin (RT) mengandalkan pertanian sebagai sumber pendapatan utama. Oleh karena

itu, perhatian terhadap kesejahteraan petani sangatlah penting, terutama karena sebagian besar rumah tangga miskin berada di wilayah pedesaan (Yacoub & Mutiaradina, 2020).

Menurut data Badan Pusat Statistik Indonesia, (2023), Jawa Barat menduduki peringkat kedua untuk kasus terbanyak kemiskinan yang ada di Indonesia. Tingkat kemiskinan di Jawa Barat pada bulan September 2022 menjadi 4,05 juta jiwa. Secara umum meskipun jumlah penduduk miskin pada Maret 2023 mencapai 3,89 juta jiwa, dibandingkan bulan September 2022 yang menurun sebanyak 182,39 ribu jiwa, Jawa Barat masih tetap menduduki peringkat kedua terbanyak untuk kasus kemiskinan yang ada di tiap-tiap Provinsi di Indonesia, maka dari itu perlunya dilakukan penanggulangan kemiskinan di Jawa Barat (BPS Jawa Barat, 2022).

Ada beberapa faktor yang menyebabkan kemiskinan di Jawa Barat yaitu tingkat pengangguran terbuka, Produk Domestik Regional Bruto (PDRB), dan kepadatan penduduk (Pusporini et al., 2023). Berdasarkan fakta tersebut pemerintah Provinsi Jawa Barat perlu untuk mengelompokkan faktor yang mempengaruhi angka

kemiskinan daerah yang ada di Jawa Barat untuk mempermudah pemerintah dalam mengatasi kemiskinan yang ada di Jawa Barat (Hidayat, 2017). Sementara itu, di Jawa Barat telah dilakukan banyak rencana untuk mengatasi kemiskinan, baik program federal maupun program khusus Pemerintah Provinsi Jawa Barat, namun efisiensi dari berbagai inisiatif tersebut belum memberikan hasil terbaik (Hidayat, 2017).

Mengelompokkan item atau variabel ke dalam kelompok-kelompok menurut ciri-ciri yang sangat sebanding dikenal sebagai analisis kluster (Putriana et al., 2016). Banyak berbagai teknik dalam analisis kluster, di antaranya adalah analisis *k-means* (Nugroho et al., 2022). Analisis kluster adalah salah satu teknik statistik yang digunakan untuk menyelesaikan proses pengelompokan. (Wijaya & Budiman, 2016). Analisis ini mengelompokkan target sehingga semua target yang serupa satu sama lain merupakan bagian dari kelompok yang sama (Wijaya & Budiman, 2016). Tujuan utama dari teknik analisis kluster adalah untuk mengelompokkan objek berdasarkan properti bersama (Yusfar et al., 2021).

Analisis kluster mempunyai berbagai cara pengelompokan. Teknik pengelompokan yang digunakan dalam penelitian ini adalah *k-means* kluster dan *average linkage*. Algoritma *k-means* merupakan metode data mining untuk mengumpulkan dan mengelompokkan bahan menjadi beberapa kategori (Harahap, 2019). Pendekatan ini dipilih karena menurut penelitian Lestari et al., (2018) metode *k-means* mengungguli teknik analisis kluster lainnya dalam hal menghasilkan temuan terbaik (Wardy et al., 2022). Kelebihan dari *k-means* yaitu sederhana, efisien dan mudah diimplementasikan (Wardy et al., 2022). Selain itu kelebihan lain dari metode *k-means* adalah kompleksitas rendah, perhitungannya cepat, dapat menangani data yang besar, anggota dalam kluster dapat disesuaikan (Widyadhana et al., 2021).

Dalam analisis kluster, ada dua jenis teknik klustering yaitu teknik *hierarki* dan teknik *non-hierarki*. *Average linkage* (Tautan rata-rata) adalah salah satu teknik pengelompokan *hierarki*. Jarak rata-rata antara setiap pasangan potensial item dalam satu kluster dan setiap objek dalam kluster lain merupakan dasar dari teknik *average linkage* (Paramadina et al., 2019). Untuk analisis

klaster *non hierarki*, metode paling sering digunakan adalah *k-means* (Widyadhana et al., 2021). Dalam penelitian ini penulis akan menggunakan metode *average linkage* dan *k-means* untuk membandingkan metode mana yang lebih akurat dalam menangani bidang kemiskinan di Jawa Barat. Kelebihan dari metode *average linkage* yaitu *dendogram* menawarkan representasi grafis dan dapat mengidentifikasi berbagai bentuk dan ukuran klaster, oleh karena itu penghitungan *average linkage* tidak perlu mengetahui berapa banyak klaster yang ada (Widyadhana et al., 2021). Karena pendekatan *average linkage* belum banyak mendapat perhatian dari para peneliti, maka hal tersebut menjadi salah satu alasan digunakannya dalam penelitian ini (Nafisah & Chandra, 2017).

Penelitian mengenai analisis klaster pada metode *k-means* klaster pernah dilakukan oleh beberapa peneliti, misalnya Lestari et al., (2018), dalam penelitiannya yang berjudul "Pengelompokan kabupaten/kota di Jawa Tengah berdasarkan tenaga kesehatan dengan menggunakan metode *ward* dan *k-means*". Temuan penelitian menunjukkan bahwa, dengan tiga klaster yang

dihasilkan, metode *k-means* adalah pendekatan yang paling efektif.

Untuk penelitian mengenai analisis kluster pada metode *average linkage* juga pernah dilakukan oleh beberapa peneliti juga, misalnya Paramadina et al. (2019), dalam penelitiannya yang berjudul "Perbandingan Analisis *Cluster* Metode *Average Linkage* dan Metode *Ward* (Kasus: IPM Provinsi Sulawesi Selatan)". Hasil penelitiannya diperoleh bahwa pengelompokkan *average linkage* dibandingkan dengan pendekatan *Ward* yang menghasilkan nilai Indeks *Dunn* sebesar 0,43, metode *average linkage* menghasilkan Indeks *Dunn* terbaik, dengan nilai 0,55 dan terbentuknya 8 kluster secara keseluruhan.

Adapun ayat yang berkaitan dengan penelitian ini tercantum dalam QS. Adz-Dzariat ayat 19-20 yang berbunyi:

﴿۱۹﴾ وَفِي أَمْوَالِهِمْ حَقٌّ لِّلسَّائِلِ وَالْمَحْرُومِ

﴿۲۰﴾ وَفِي الْأَرْضِ آيَاتٌ لِّلْمُؤْمِنِينَ

Artinya:

“Pada harta benda mereka ada hak bagi orang miskin yang meminta dan yang tidak meminta (19). Di bumi terdapat tanda-tanda (kebesaran Allah) bagi orang-orang yang yakin (20)”. (QS. Adz Dzariat ayat 19-20).

Maksud dari ayat di atas adalah dari harta yang kita miliki ada sebagian hak untuk orang miskin, dengan ciptaan Allah yang berbagai macam di dunia ini itulah tanda kekuasaan Allah sesungguhnya (Quran.nu.or.id, 2023). Dari ayat di atas terlihat bahwa kemiskinan merupakan hal yang sangat penting untuk diatasi.

Berdasarkan uraian di atas, metode *k-means* dan *average linkage* mempunyai beberapa kelebihan khususnya, bahwa hal ini lebih sederhana dan memakan waktu lebih sedikit untuk melaksanakannya, dan bahwa temuan-temuan penilaiannya juga cukup baik. Mengingat permasalahan di atas, penulis berkeinginan untuk melakukan penelitian metode *k-means* dan *average linkage* terkait pengelompokkan kemiskinan di Provinsi Jawa Barat yang dirasa bermanfaat untuk mempermudah pemerintah dalam melihat kebutuhan di tiap-tiap daerah di Provinsi Jawa Barat untuk mengoptimalkan

penanggulangan kemiskinan di Provinsi Jawa Barat. Oleh karena itu, penulis tertarik untuk melakukan penelitian ini dengan judul "PERBANDINGAN ANALISIS KLASSTER *K-MEANS* DAN *AVERAGE LINKAGE* PADA PENGELOMPOKKAN KEMISKINAN DI PROVINSI JAWA BARAT PADA TAHUN 2022".

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, maka permasalahan yang dapat diangkat sebagai berikut:

- a. Bagaimana pengelompokkan kemiskinan di Provinsi Jawa Barat pada tahun 2022 menggunakan metode *k-means* dan *average linkage*?
- b. Bagaimana gambaran karakteristik dari masing-masing klaster hasil dari metode *k-means* dan *average linkage* berdasarkan pengelompokkan kemiskinan di Provinsi Jawa Barat pada tahun 2022?
- c. Bagaimana perbandingan hasil pengelompokkan kemiskinan di Provinsi Jawa Barat pada tahun 2022 antara metode *k-means*

dan *average linkage* dalam menentukan metode terbaik?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang diangkat, maka tujuan yang ingin dicapai dalam penelitian ini adalah berikut:

- a. Untuk mengetahui pengelompokkan kemiskinan di Provinsi Jawa Barat pada tahun 2022 menggunakan metode *k-means* dan *average linkage*.
- b. Untuk mengetahui gambaran karakteristik masing-masing klaster hasil dari metode *k-means* dan *average linkage* pada kemiskinan di Provinsi Jawa Barat pada tahun 2022.
- c. Untuk mengetahui perbandingan hasil pengelompokkan kemiskinan di Provinsi Jawa Barat pada tahun 2022 antara metode *k-means* dan *average linkage* dalam penentuan metode terbaik.

1.4 Manfaat Penelitian

Dengan tercapainya tujuan penelitian, diharapkan dapat memberikan manfaat baik bagi civitas akademik UIN Walisongo Semarang, bagi penulis dan bagi khalayak umum, yaitu:

a. Bagi Penulis

Penelitian ini diharapkan dapat memberikan pengetahuan tentang cara menganalisis kemiskinan di Provinsi Jawa Barat dengan metode dalam analisis kluster.

b. Bagi Pembaca

Dapat digunakan sebagai bahan referensi untuk penelitian lebih lanjut.

c. Bagi Universitas

Dapat dilakukan pengembangan ilmu pengetahuan dalam bidang statistika khususnya dengan analisis kluster.

d. Bagi Pemerintah

Sebagai masukan dalam membuat suatu kebijakan memberantas kemiskinan daerah di Provinsi Jawa Barat.

1.5 Batasan Masalah

Pada penelitian ini batasan masalah yang digunakan adalah:

1. Metode analisis kluster yang digunakan adalah metode *k-means* dan *average linkage*.
2. Pemilihan metode terbaik dari metode *kmeans* dan *average linkage* dengan menghitung nilai *silhouette coefficient*..
3. Data yang digunakan adalah data kemiskinan di Provinsi Jawa Barat dengan menggunakan empat variabel yaitu garis kemiskinan, jumlah penduduk miskin, jumlah penduduk miskin ekstrem, dan pengeluaran perkapita.

BAB II

LANDASAN PUSTAKA

2.1 Tinjauan Pustaka

2.1.1 Statistik Deskriptif

Statistik deskriptif adalah statistik yang berguna untuk mendeskripsikan atau menganalisis data penelitian, tetapi bukan untuk menarik kesimpulan lebih lanjut tentang karakteristik populasi. Ruang lingkup statistik deskriptif meliputi: konsep statistika dasar, distribusi frekuensi, pengukuran titik pusat, pengukuran sebaran, *skewness* dan *kurtosis*, penyajian data dalam bentuk diagram grafik (grafik batang, diagram garis, *histogram* batang, *poligon*, *ogive*), angka indeks, dan deret waktu. Secara umum fungsi statistik deskriptif adalah mengumpulkan data, mengolah data, menyajikan data, dan menganalisis data (Farisi et al., 2023). Analisis statistik deskriptif merupakan bagian dalam statistika tentang pengumpulan data, penyajian,

penentuan nilai-nilai statistika, pembuatan diagram atau gambar tentang suatu hal, dimana data yang disajikan dalam bentuk yang lebih mudah dipahami atau dibaca (Pramesthy et al., 2021).

Statistik deskriptif merupakan pengukuran yang menggambarkan atau memberikan garis besar subjek yang diteliti melalui contoh atau informasi penduduk tanpa penyelidikan atau membuat kesimpulan yang berlaku untuk orang-orang pada umumnya (Agama & Karawang, 2010). Sedangkan menurut Yuhelmi et al.,(2018), statistik deskriptif cabang merupakan ilmu yang mempelajari pengumpulan data dan penyajiannya secara komprehensif. Satu-satunya tujuan statistik deskriptif adalah untuk mengkarakterisasi atau memberikan rincian tentang data atau keadaan.

Berdasarkan pengertian diatas, maka dapat disimpulkan bahwa statistik deskriptif metode yang digunakan untuk menggambarkan suatu kondisi, gejala, atau masalah tanpa melakukan analisis dengan penarikan kesimpulan atau

menarik kesimpulan tentang statistik deskriptif mengenai kumpulan data yang tersedia saja.

2.1.2 Analisis Multivariat

Data dari dua variabel atau lebih dapat dianalisis secara bersamaan menggunakan jenis analisis statistik yang disebut analisis multivariat. Analisis *depedensi* dan analisis *interdependensi* adalah dua kategori besar yang termasuk dalam teknik analisis multivariat. Suatu variabel terikat dapat diperkirakan atau dijelaskan dengan menggunakan analisis *depedensi* apabila terdapat dua atau lebih variabel bebas. Analisis regresi berganda, analisis diskriminan, analisis konjoin, analisis varians multivariat, dan analisis korelasi *kanonik* adalah contoh analisis *depedensi*. Analisis *interdependensi*, di sisi lain, digunakan untuk mengelompokkan variabel atau memberikan signifikansi pada beberapa variabel. Analisis faktor, analisis klaster, dan penskalaan multidimensi adalah contoh analisis saling *depedensi* (Wijaya & Budiman, 2016).

Dalam analisis multivariat, tipe data dikategorikan menjadi lima tipe atau skala menurut (Bilson Simamora, 2005), berdasarkan cara pengumpulan atau pengumpulan data. Ini adalah sebagai berikut:

1. Skala Nominal

Sumber data nominal adalah *observasional*. Angka-angka yang digunakan bersifat *diskontinu* dan tidak mengenal urutan jika datanya direpresentasikan sebagai data numerik (kuantitatif). Ini menyiratkan bahwa terlepas dari ukuran atau penempatannya, setiap elemen tidak ada artinya.

2. Skala Ordinal

Data ordinal berasal dari pertanyaan berskala suatu variabel atau dari observasi. Observasi menghasilkan data kualitatif sebagai hasilnya. Ketika data direpresentasikan secara numerik, angka-angka tersebut bersifat *diskrit* dan disusun dalam *hierarki* berdasarkan kualitas-kualitas yang terkait.

3. Skala Kardinal

Output dari penghitungan atau komputasi suatu variabel menghasilkan data utama. Bilangan *diskrit* kuantitatif, biasanya dinyatakan sebagai bilangan pokok.

4. Skala Interval

Data interval berasal dari hasil mengukur suatu variabel. Data diasumsikan berbentuk bilangan kontinu mempunyai ukuran urutan, seperti data dengan ordinal. Pada skala interval tidak memiliki nol mutlak, artinya jika suatu responden variabelnya bernilai nol bukan berarti tidak memiliki substansi sama sekali.

5. Skala Rasio

Hasil pengukuran suatu variabel menghasilkan data rasio. Diasumsikan bahwa data berbentuk bilangan kontinu yang hampir identik dengan skala interval dengan nilai nol sebagai pembedanya. Suatu variabel respon tidak mempunyai substansi sama sekali jika nilainya nol pada skala rasio yang mempunyai nilai nol mutlak. (Sukestiyarno .,et.al, 2016).

2.1.3 Data Mining

Data mining didefinisikan sebagai penambangan data, menurut Prahartiwi & Dari (2019) yaitu “penggalian seperti pengetahuan dari perkumpulan data yang besar”. Beberapa pendapat orang mengibaratkan data mining adalah sebutan lain dari *Knowledge Discovery from Data* (KDD) (Mulya, 2019). Berdasarkan beberapa definisi tersebut dapat disimpulkan bahwa data mining merupakan suatu pengetahuan yang mencari informasi penting dan berharga yang tersembunyi dan belum diketahui pada suatu database yang sangat besar sehingga akan ditemukan suatu pola menarik yang sebelumnya belum diketahui.

Dalam data mining terdapat tahap-tahap yang penting dalam proses *knowledge discovery* yang terdiri dari 7 tahapan sebagai berikut:

- 1) Pembersihan data (*Data cleaning*)

Data cleaning atau pembersihan data adalah proses menghapus data duplikat, memeriksa ketidak konsistenan

data, dan memperbaiki kesalahan pada data, seperti kesalahan ketik. Data yang diperoleh dari hasil pengujian atau database perusahaan biasanya mengandung kesalahan, seperti kesalahan ketik atau data yang hilang atau salah.

- 2) Integrasi data (*Data integration*) Integrasi data merupakan proses melengkapi data terkini dengan informasi lebih lanjut atau data yang relevan dapat dibahas sebagai konsolidasi data dari database yang berbeda menjadi satu data baru yang dibutuhkan oleh *knowledge discovery* atau penemuan pengetahuan.
- 3) Pemilihan data (*Data selection*) Pemilihan dan analisis data aktivitas terkait. Data terkait pemilu disimpan di database berbeda.
- 4) Transformasi data (*Data transformation*) Transformasi data merupakan proses transformasi data ke dalam format tertentu agar sesuai untuk prosedur penambangan

data. Teknik umum tertentu, seperti analisis asosiasi dan pengelompokan, misalnya, terbatas pada penerimaan data berupa kategorik.

- 5) Data mining
Proses penerapan strategi, taktik, atau algoritma tertentu untuk mengidentifikasi pola atau informasi yang menarik.
- 6) Evaluasi pola (*Pattern Evaluation*)
Tentukan pola yang sangat menarik dari hasil data mining. Temuan teknik data mining pada langkah ini adalah model yang tipikal dan model prediktif dievaluasi untuk menentukan apakah hipotesis yang ada benar-benar terpenuhi
- 7) Penyajian pengetahuan (*Knowledge Presentation*)
Penyajian pengetahuan merupakan proses yang menampilkan pola atau model proses penambangan data, dan alat bantu visualisasinya dalam mengkomunikasikan temuan penambangan data dengan cara

yang dapat dipahami (Prahartiwi & Dari, 2019).

2.1.4 Standarisasi Data

Salah satu proses yang harus dilakukan dalam proses analisis yaitu standarisasi data. Proses standarisasi data yaitu untuk menyamakan menjadi satuan data. Dari persamaan satuan data maka akan diperoleh perhitungan jarak menjadi valid (Goejantoro & Desi Yuniarti, 2019). Perlunya melakukan proses standarisasi data untuk menentukan variabel yang diteliti menjadikan sebagai satu ukuran satuan (Zulyanti & Noeryanti, 2022). Menemukan nilai *mean* dan *varians* setiap variabel adalah cara mendapatkan nilai standarisasi, berikut rumus standarisasi:

$$\bar{x}_k = \frac{1}{N} \sum_{i=1}^N x_{ik} \quad (2.1)$$

$$\sigma_k^2 = \frac{1}{N-1} \sum_{i=1}^N (x_{ik} - \bar{x}_k)^2 \quad (2.2)$$

$$\hat{x}_{ik} = \frac{x_{ik} - \bar{x}_i}{\sigma^2} \quad (2.3)$$

dimana:

- N : jumlah data
- \bar{x}_{ik} : data ke- i variabel ke- k dimana $k = 1, 2, \dots, n$
- \bar{x}_i : rata-rata variabel ke- k
- σ_k : simpangan baku variabel ke- k
- σ_{k^2} : variansi variabel ke- k
- \bar{x}_{ik} : standarisasi data i ke variabel k
(Suyanto et al., 2021).

2.1.5 Jarak *Euclidean*

Jarak *Euclidean* adalah pengukuran untuk menghitung jarak dua titik dalam *euclidean* digunakan untuk menghitung jarak dalam satu dimensi antara dua titik, menghasilkan hasil yang mirip dengan perhitungan *Pythagoras*. Persamaan jarak *euclidean* adalah sebagai berikut:

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (2.4)$$

Keterangan :

d = Jarak

\bar{x}_1 = Koordinat *latitude* 1

\bar{x}_2 = Koordinat *latitude* 2

y_1 = Koordinat *longitude* 1

y_2 = Koordinat *longitude* 2 (Palupi et al., 2021).

2.1.6 Analisis Klaster

Analisis klaster merupakan teknik multivariat dengan tujuan, pengelompokan, dimana masing-masing kelompok mempunyai ciri-ciri yang relatif sama (homogen), namun ciriciri kelompoknya berbeda satu sama lain (Wulan Permata Sari & Tata Sutabri, 2023). Secara umum, suatu objek ditempatkan dalam suatu klaster atau kelompok yang di dalamnya terdapat kemungkinan lebih tinggi adanya hubungan (korelasi) antara objek tersebut dengan hal-hal lain dalam klaster tersebut dibandingkan dengan objek lain. Ada atau tidaknya hubungan yang kuat antar

objek menentukan terbentuknya suatu kluster (Wulan Permata Sari & Tata Sutabri, 2023).

Analisis kluster Variabel *independen* dan *dependen* dalam analisis multivariat pendekatan *interdependensi* tidak dibedakan satu sama lain. Dengan kata lain, analisis kluster adalah serangkaian teknik statistik yang mengidentifikasi pengelompokan sampel berdasarkan atribut bersama. Item dengan *heterogenitas* tinggi antar kluster dan *homogenitas* tinggi dalam kluster dikelompokkan dalam analisis ini (Wahyuni et al., 2013). Terdapat dua asumsi dalam analisis kluster menurut (Ningrat et al., 2016) yaitu:

1. Asumsi Korelasi

Korelasi adalah sebuah nilai yang digunakan untuk mengukur derajat keeratan dari hubungan antara dua variabel. Untuk mengetahui hubungan yang positif dan signifikan antara variabel-variabel *independent* dengan variabel *dependen* (S. W. K. Dewi et al., 2021).

Tisantri et al., (2019) mengungkapkan dalam statistik Hadyan parametrik, koefisien

korelasi yang paling sering digunakan adalah koefisien korelasi momen hasil kali *Pearson* yang dinyatakan dengan r , dengan rumus r sebagai berikut:

$$r_{\bar{x}\bar{y}} = \frac{(\sum(X-\bar{X})(Y-\bar{Y}))}{[\sum(X-\bar{X})^2(\sum(Y-\bar{Y})^2)]^{\frac{1}{2}}} \quad (2.5)$$

Keterangan:

$r_{\bar{x}\bar{y}}$ = nilai koefisien korelasi

X = data 1

X = data 2

\bar{X} = rata rata dari data X

\bar{Y} = rata rata dari data Y (Hadyan Tisantri et al., 2019)

Jika koefisien korelasi berada di antara -1 dan 1, maka dapat bernilai positif atau negatif. Koefisien korelasi negatif menunjukkan koefisien negatif, dan sebaliknya (Sihombing & Bangun, 2019). Interpretasi terhadap koefisien korelasi ditunjukkan pada **Tabel 2.1**.

Tabel 2.1 Interval Kategori Hubungan Korelasi

Besar Koefisien Korelasi (positif atau negatif)	Interpretasi Koefisien Korelasi
0,80-1	Sangat kuat
0,60-0,799	Kuat
0,40-0,599	Cukup kuat
0,20-0,399	Rendah
0-0,199	Sangat rendah

(Sihombing & Bangun, 2019).

2. Asumsi *Representative*

Asumsi *representative* dapat dinyatakan dengan Tes *Kaiser Mayer Olkin* (KMO). Baik kecukupan pengambilan sampel secara keseluruhan maupun kecukupan pengambilan sampel untuk setiap indikator dinilai dengan menggunakan uji KMO. Uji KMO digunakan untuk mengetahui validitas faktor penelitian. Persamaan 2.6 menyajikan rumus uji KMO sebagai berikut.

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (2.6)$$

$$a_{ij} = \frac{-r_{ij}}{\sqrt{r_{ij}r_{ij}}} \quad (2.7)$$

dimana:

p = banyaknya variabel

r_{ij} = koefisien korelasi antara variabel i dan j

a_{ij} = koefisien korelasi parsial antara variabel i dan j

Asumsi representative terpenuhi dan uji analisis selanjutnya dapat dilakukan jika nilai KMO lebih besar dari 0,5 (Ningrat et al., 2016).

2.1.7 Algoritma *K-means*

K-means kluster adalah pembentukan satu atau lebih kluster atau kelompok objek dengan menggunakan salah satu metode analisis kluster non-hierarki; dengan cara ini, objek-objek dengan karakteristik serupa dikelompokkan dalam satu kluster, dan objek-objek dengan karakteristik berbeda dikelompokkan dalam kluster lain. Satu lagi (Ediyanto et al., 2013). Tujuan

pengelompokan yaitu untuk mengurangi fungsi tujuan proses pengelompokan, yang pada dasarnya berupaya meningkatkan variasi antar kluster dan meminimalkan variasi dalam satu kluster (Bangun, 2016). Algoritma *k-means* bertujuan untuk menggabungkan seluruh data yang ada ke dalam satu kelompok, dimana setiap data kelompok memiliki atribut yang unik dari data kelompok lainnya (Halilintar R & Farina Nur I, 2018).

K-means umum digunakan dengan data mining (Widyadhana et al., 2021). Sebagai contoh $X = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n\}$ adalah data yang akan dianalisis dan $V = \{v_1, v_2, \dots, v_n\}$ adalah titik pusat grup data X dalam dimensi (\mathbb{R}^p). Dalam contoh ini, angka n , p dan kluster masing-masing mewakili jumlah objek, variabel, dan partisi atau pengelompokan (Widyadhana et al., 2021). Rumus berikut dapat digunakan untuk menentukan titik pusat kluster.

$$v_{ij} = \frac{1}{n} \sum_{k=1}^{n_i} x_{kj} \quad (2.8)$$

dimana n_i adalah banyaknya objek pada kelompok ke- i , \bar{x}_{kj} adalah nilai pengamatan dari objek k ke variabel j , dan v_{ij} adalah titik pusat kelompok ke- i pada variabel j . Ini adalah salah satu cara untuk menulis tentang pengelompokan data.

$$v_{ij} = \begin{cases} 1, d = \min (d^2_{ik}(x_k, v_i)) \\ 0, d \text{ yang lainnya} \end{cases} \quad (2.9)$$

dimana \bar{x}_k adalah objek pengamatan ke- k , v_i adalah nilai rata-rata pusat kelompok i , dan μ_{ik} adalah nilai k anggota objek pada kluster i . Pengelompokan ini bertujuan untuk memaksimalkan keberagaman antar kelompok sekaligus memupuk keberagaman di dalam kelompok. Dengan kata lain, tujuan pengelompokan ini adalah untuk mengurangi fungsi tujuan (Cebeci & Yildiz, 2015). Fungsi tujuan ini dapat dinyatakan secara matematis dengan menggunakan rumus berikut.

$$J(X, V) = \sum_{k=1}^{n_i} \sum_{i=1}^c \mu_{ik} d^2_{ik} \quad (2.10)$$

dimana c adalah jumlah kluster, d_{ik}^2 adalah jarak *euclidean* antara suatu objek dengan pusat kluster ke- i , dan n_i adalah jumlah objek yang terdapat pada grup ke- i (Widyadhana et al., 2021).

Pendekatan *k-means* kluster digunakan dalam penelitian ini untuk mengkategorikan penyebab apa saja yang menjadi permasalahan utama kemiskinan di Provinsi Jawa Barat. Ada banyak faktor yang perlu diukur seperti garis kemiskinan, jumlah penduduk miskin, jumlah penduduk miskin ekstrem, dan pengeluaran per kapita.

2.1.8 Silhouette Coefficient

Silhouette Saat mengklasifikasikan objek, *coefficient* adalah metrik penentu. Jika koefisien yang diperoleh mencapai nilai 1 maka kluster yang dihasilkan dianggap dapat diterima, jika koefisiennya mencapai nilai -1 maka dianggap kurang baik. (Suraya & Wijayanto, 2022). Metode *silhouette coefficient* merupakan gabungan dari metode *separation* dan *cohesion*. Rumus

persamaan *silhouette coefficient* yaitu (Handoyo et al., 2014):

1. Jarak rata-rata dari data dihitung dengan menggunakan rumus yaitu:

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.11)$$

dimana:

$a(i)$: rata-rata objek (i) ke semua objek lain di kluster yang sama

$d(i, j)$: jarak antara data i dan j

A : banyaknya data

i : anggota data i

j : anggota data j

2. rumus jarak rata-rata terendah antara data I dengan jumlah total data pada kluster lain yaitu:

$$b(i) = \min_{C \neq A} d(i, j) \quad (2.12)$$

dimana:

$b(i)$: nilai minimum rata-rata jarak data (i)

Persamaan *silhouette coefficient*:

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (2.13)$$

Tabel berikut menampilkan kriteria untuk menentukan baik atau tidaknya hasil kluster dengan menggunakan *silhouette coefficient* versi Kaufman dan Rousseeuw (Qadrini, 2020).

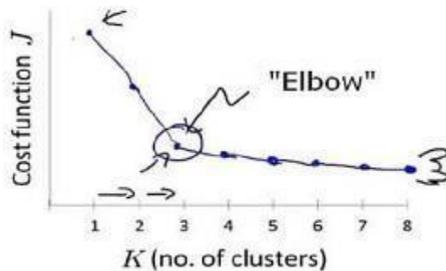
Tabel 2.2 Interpretasi Nilai *Silhouette Coefficient*

Nilai <i>Silhouette Index</i>	Keterangan
0,71-1,00	Kuat
0,51-0,70	Baik
0,26-0,50	Lemah
$\leq 0,25$	Tidak Terstruktur

2.1.9 Metode *Elbow*

Metode *elbow* membandingkan 35 persen jumlah kluster yang membentuk *elbow* pada posisi tertentu untuk menentukan jumlah kluster yang ideal. Representasi grafis dari hasil persentase untuk setiap kluster disediakan. Nilai kluster pertama dikatakan terbaik jika nilainya membentuk sudut pada grafik dengan nilai

kluster kedua, atau jika nilainya turun paling besar. Ada beberapa nilai K pada **Gambar 2.1** yang mengalami penurunan terbesar, dan nilai K tersebut kemudian turun secara bertahap hingga nilai K akhir stabil. Misalnya pada **Gambar 2.1** dapat dilihat bahwa kluster bernilai $K=2$ hingga $K=3$ dan $K=3$ hingga $K=4$ (Muningsih & Kiswati, 2018).



Gambar 2.1 Gambaran Analisis Uji Metode

Elbow

Penjelasan **Gambar 2.1** ideal selama titik penurunannya membentuk siku (Aditya et al., 2020). Meskipun demikian, ada situasi di mana nilai jatuh yang sangat kecil membuat sulit untuk membedakan bentuk siku. Jadi, ada situasi

di mana siku tidak ada atau ada tetapi sedikit berbeda. (Jollyta et al., 2021).

2.1.10 Metode *Gap Statistic*

Gap statistic adalah metode untuk menghitung kebaikan ukuran pengelompokan dengan melihat nilai $\log W$ dengan $\log E$ untuk setiap jumlah kluster (Dewa & Jatipaningrum, 2019). *Gap statistic* akan menentukan jumlah kluster terbentuk dengan mengamati dalam grafik *gap statistic*. Berikut ini adalah statistik *gap statistic* untuk nilai tertentu:

$$Gap_n(k) = E_n^* \log(W_k) - \log(W_k) \quad (2.14)$$

Keterangan:

E_n^* = ukuran sampel n

(W_k) = log rata-rata (Dewa & Jatipaningrum, 2019).

2.1.11 Metode *Hierarki*

Tipe dasar dalam metode *hierarki* bisa *agglomeratif* atau *defisif*. Pada pengklasteran *agglomeratif*, Dalam pengelompokan defisit, semua objek awalnya dikelompokkan ke dalam

satu klaster, namun dalam kasus sebelumnya, objek ditempatkan terlebih dahulu dalam kelompok yang lebih kecil dan kemudian secara bertahap dikelompokkan ke dalam klaster yang lebih besar. Barang-barang tersebut kemudian secara bertahap dibagi ke dalam berbagai kelompok. Ada dua, tiga, dan seterusnya klaster (Goreti et al., 2016). Ada 5 metode *hierarki agglomeratif* dalam pembentukan klaster yaitu:

- a. Pautan tunggal (*single linkage*)
- b. Pautan lengkap (*complete linkage*)
- c. Pautan rata-rata (*average linkage*)
- d. Metode *Ward* (*Ward's method*)
- e. Metode pusat (*centroid*) (Paramadina et al., 2019).

2.1.12 Algoritma *Average Linkage*

Jarak rata-rata antara setiap pasangan potensial item dalam satu klaster dengan seluruh objek dalam klaster lainnya merupakan dasar dari metode *average linkage*. Jarak antara dua klaster, atau jarak rata-rata, ditentukan oleh hubungan rata-

rata dan dihitung untuk setiap kluster secara individual. Langkah pertama dalam proses *average linkage* adalah mendefinisikan matriks $D = \{d_{ik}\}$ untuk mencari objek terdekat, seperti U dan V . Setelah itu, item-item tersebut digabungkan menjadi sebuah kluster (UV) . Selanjutnya pemisahan dari kluster lain (W) dan (UV) (Paramadina et al., 2019).

$$d((UV)W) = \frac{\sum_i \sum_k d_{ik}}{N_{UV}N_W} \quad (2.15)$$

Dimana:

d_{ik} = jarak antar objek i dalam kluster (UV)

dan objek k dalam kluster w

N_{UV} = jumlah item pada kluster UV

N_W = jumlah item pada kluster W (Paramadina et al., 2019).

2.1.13 Pengertian Kemiskinan

Menurut kamus umum bahasa Indonesia, istilah “miskin” mengacu pada tidak mempunyai harta benda dan berada dalam kemiskinan

(berpenghasilan rendah). Dalam hal ini apa yang dimaksud dengan kemiskinan adalah orang yang ditenangkan oleh kefakiran dan ia adalah orang yang memiliki sesuatu yang tidak mencukupi kebutuhannya (Cahya, 2015). Kemiskinan merupakan refleksi dari ketidakmampuan seseorang untuk memenuhi kebutuhannya sesuai dengan standar yang berlaku (Mulia, 2022).

Dalam bukunya Pembangunan untuk Rakyat, Ginanjar Kartasasmita menjelaskan penyebab kemiskinan setidaknya ada empat, yaitu sebagai berikut:

1. Rendahnya taraf pendidikan Rendahnya tingkat pendidikan terampil membatasi kemampuan seseorang untuk mengeksplorasi diri dan menyulitkan dalam mencari pekerjaan.
2. Rendahnya derajat kesehatan kondisi fisik yang tidak memadai, fungsi kognitif, dan pembentukan ide adalah konsekuensi dari kesehatan dan gizi yang tidak memadai.

3. Terbatasnya lapangan kerja Kondisi kemiskinan yang dipengaruhi kesehatan dan pendidikan dipersulit dengan lapangan kerja yang sangat terbatas. Jika terdapat lapangan kerja maka potensi untuk memotong rantai kemiskinan.
4. Kondisi keterisolasian Banyak masyarakat miskin yang tidak mampu memenuhi kebutuhan hidupnya karena terjerat dan terlindungi. Akses kesehatan, pendidikan, dan layanan masyarakat penting lainnya akan lebih sulit dilakukan jika seseorang tinggal di daerah terpencil (V. E. Dewi, 2021).

2.1.14 Ciri Kemiskinan

Kemiskinan didefinisikan sebagai ketidakmampuan untuk membayar kebutuhan seperti makanan dan non-makanan. Jika rata-rata pengeluaran per kapita bulanan penduduk kurang dari garis kemiskinan, maka mereka tergolong miskin (Bps.go.id, 2023). Bank Dunia dan PBB mendefinisikan kemiskinan parah sebagai

mempunyai PPP (*purchasing power parity*) harian sebesar 1,9 dolar AS. Pada tahun 2021, PPP 1,9 dolar AS setara dengan sekitar 11.941 rupiah jika dijabarkan ke rupiah (bps.go.id, 2023).

Garis Kemiskinan (GK) mewakili jumlah kebutuhan uang dalam rupiah paling sedikit yang harus dikeluarkan seseorang untuk memenuhi kebutuhan pokok baik makanan maupun non makanan selama sebulan. Jumlah paling sedikit yang dibutuhkan untuk memenuhi kebutuhan pangan seseorang dikenal dengan Garis Kemiskinan Pangan (GKM), yaitu sebesar 2100 kalori per orang per hari. Garis Kemiskinan Non-Makanan (GKNM) mewakili jumlah terendah uang yang dibutuhkan untuk memenuhi kebutuhan non-makanan, seperti perumahan, pakaian, layanan kesehatan, dan pendidikan (bps.go.id, 2023). Pengeluaran rata-rata bulanan per kapita dihitung dengan membagi total biaya konsumsi bulanan yang dikeluarkan oleh seluruh anggota rumah tangga baik melalui pembelian, hadiah, atau produksi sendiri dengan

jumlah total orang yang ada di rumah (bps.go.id, 2023).

2.2 Penelitian Terdahulu

Penelitian terdahulu sangat penting bagi penulis untuk memiliki pilihan menuju pada penelitian baru, agar dapat memiliki pilihan untuk mengetahui hubungan antara penelitian terdahulu dengan penelitian baru yang dilakukan oleh penulis, untuk mencegah pencurian sastra, untuk dapat berkontribusi pada penelitian ini terhadap perkembangan ilmu pengetahuan. Penelitian dengan menggunakan analisis *k-means* kluster dan *average linkage* untuk mengelompokkan secara umumnya pernah dilakukan oleh beberapa peneliti terdahulu. Ada beberapa penelitian yang berkaitan dengan penelitian ini sebagai berikut.

- a. Lestari et al., (2018), dalam penelitiannya yang berjudul "Pengelompokkan Kabupaten/Kota di Jawa Tengah Berdasarkan Tenaga Kesehatan dengan Menggunakan Metode *Ward* dan *K-Means*"

pada jurnal *Fourier*. Tiga kluster diidentifikasi berdasarkan temuan penelitian, dan pendekatan *k-means* yang memiliki rasio deviasi standar sebesar 0,2974% ditentukan sebagai pendekatan yang paling efektif. Persamaannya, metode yang digunakan sama yaitu *k-means*. Perbedaannya, pada penelitian yang akan dilakukan menggunakan metode *k-means* dan *average linkage*.

- b. Putriana et al., (2016), dalam penelitiannya yang berjudul "Metode *Cluster Analysis* untuk Pengelompokkan Kabupaten di Provinsi Jawa Tengah Berdasarkan Variabel yang Mempengaruhi Kemiskinan Pada Tahun 2013" dalam Jurnal Statistik Industri dan Komputasi dengan tujuan untuk mengklasifikasikan kota dan kabupaten di Jawa Tengah menurut variabel yang mempengaruhi tingkat kemiskinan. Temuan penelitian ini menunjukkan bahwa, dari tiga kelompok, pendekatan *k-means*

adalah yang paling efektif. Nilai RMSSTD dan RS digunakan untuk menentukan kluster mana yang terbaik. Persamaannya, metode yang digunakan sama yaitu pada metode *kmeans*. Perbedaannya yaitu pada variabelvariabel yang digunakan.

- c. Handoko et al., (2020), dalam penelitiannya yang berjudul "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode *K-Means Clustering*" pada Jurnal Ilmiah Teknologi dan Rekayasa yang memiliki tujuan mengembangkan perangkat lunak yang menggunakan algoritma *k-means* untuk mengklasifikasikan data transaksi penjualan paket data Telkomsel ke dalam kelompok regional dengan penjualan tinggi, rendah, dan rendah. Teknik pengelompokan. Persamaannya, terletak pada metode *k-means* kluster untuk menentukan metode pengelompokan

variabel-variabelnya. Perbedaanya, penelitian ini menggunakan metode *kmeans* klaster berdasarkan data kemiskinan penduduk di Provinsi Jawa Barat.

2.3 Kerangka Berpikir

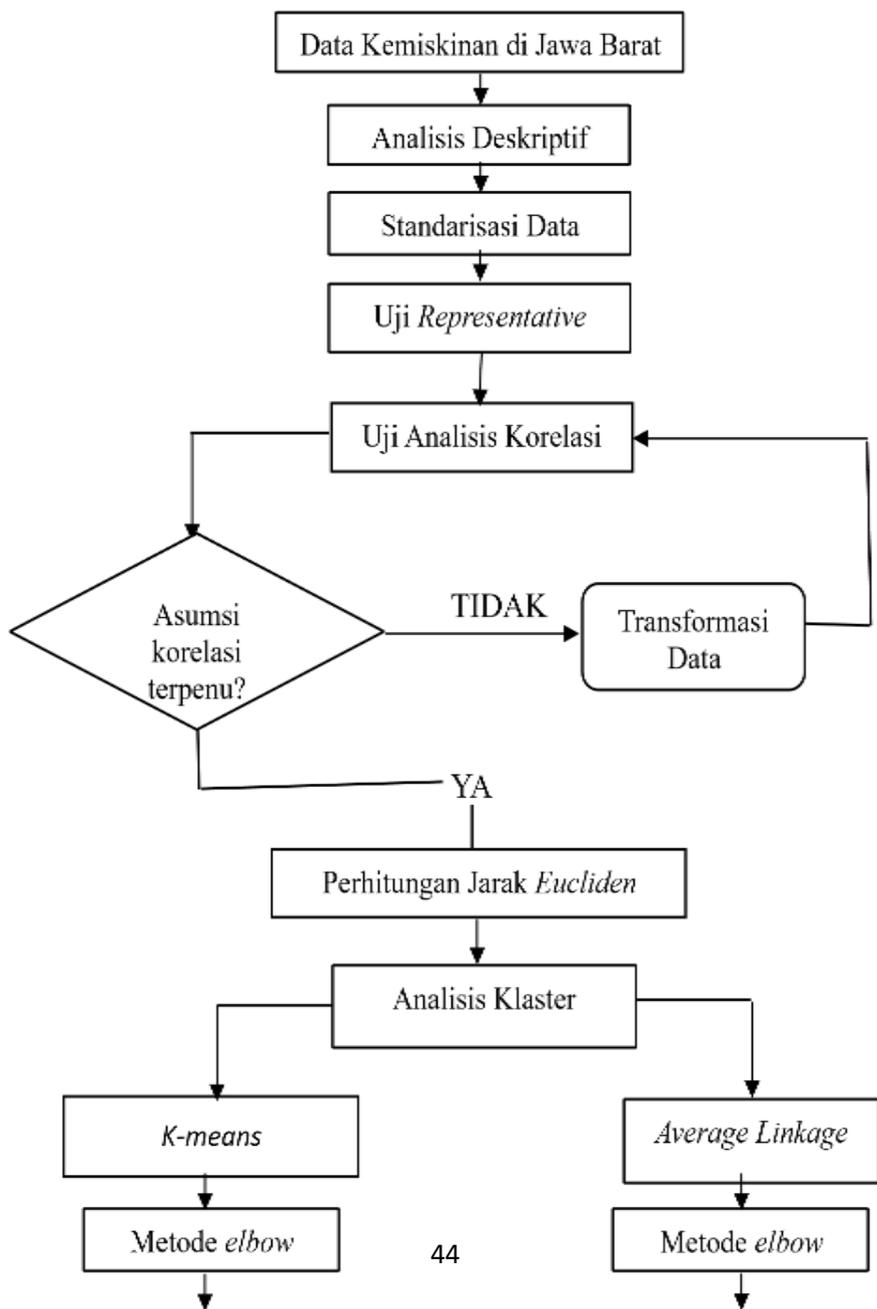
Setiap provinsi pasti mempunyai banyak permasalahan termasuk dengan kemiskinan. Untuk mengetahui penyebab utama dari kemiskinan, maka perlu untuk mengelompokkan. Karena analisis klaster dapat mengelompokkan item dengan *homogenitas* tinggi di dalam klaster dan *heterogenitas* tinggi di seluruh klaster, maka analisis ini cocok digunakan untuk pengelompokan. Metode analisis *k-means* dan *average linkage*, keduanya telah diuji dalam penelitian sebelumnya, dan hasilnya menunjukkan hasil yang efektif.

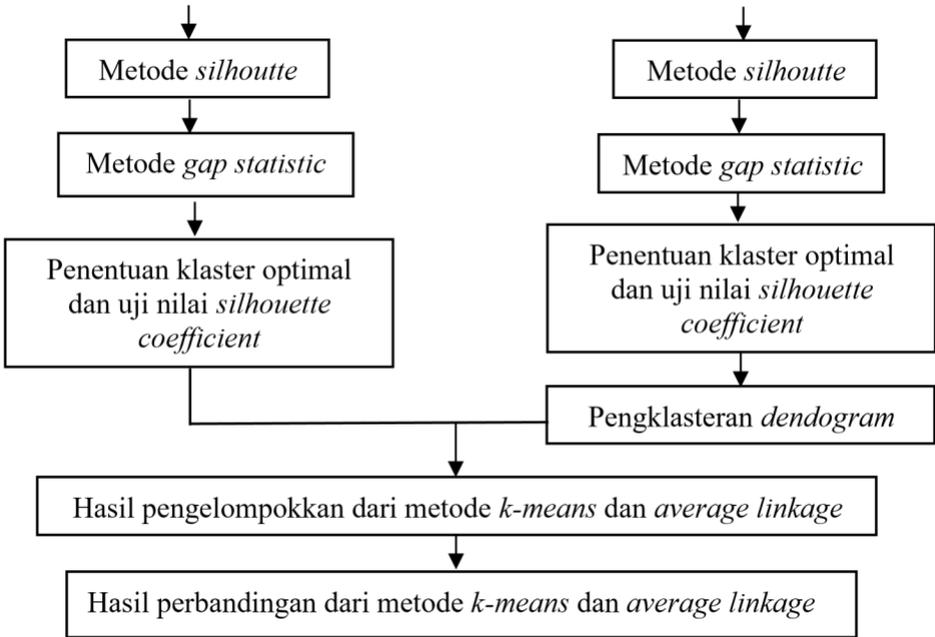
Data angka kemiskinan di setiap kabupaten dan kota di Jawa Barat dikumpulkan sebagai langkah awal dalam penelitian ini.

Selanjutnya, konfirmasi faktor-faktor penting yang terkait dengan isu yang sedang dibahas.

Selanjutnya melakukan analisis deskriptif untuk mendapatkan pemahaman umum kemiskinan penduduk di Jawa Barat. Tahap selanjutnya melakukan standarisasi data untuk menyamakan variabel yang berbeda, kemudian melakukan uji korelasi dan uji *representative*. Setelah itu, melakukan perhitungan jarak *euclidean*. Dilanjutkan ada proses pengelompokan data dengan metode *k-means* dan *average linkage*. Selanjutnya, temuan pengelompokan akan dibandingkan untuk menentukan pendekatan mana yang paling efektif untuk penelitian ini

Bagan berikut ini disajikan untuk membantu memvisualisasikan tahapan penelitian berdasarkan kerangka teori dasar penelitian yang telah dikemukakan sebelumnya:





Gambar 2.2 Tahapan Penelitian

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan metodologi penelitian kuantitatif. Penelitian kuantitatif melibatkan penggunaan data dalam bentuk angka sebagai alat untuk menemukan pengetahuan tentang topik yang diminati (Rahmadi, 2011). Dengan mengedepankan konsep, teori, atau ide, penelitian ini memberikan solusi terhadap permasalahan yang muncul dalam kehidupan sehari-hari. Jenis penelitian ini akan memberikan gambaran bagaimana pemerintah mengatasi kemiskinan di Provinsi Jawa Barat.

3.2 Populasi dan Sampel Penelitian

Populasi subjek penelitian ini adalah Kabupaten/Kota di Provinsi Jawa Barat yang berjumlah 27 diantaranya ada 18 Kabupaten dan 9 Kota. Teknik sampel yang digunakan dalam

penelitian ini adalah Teknik sampling jenuh karena jumlah datanya yaitu kabupaten/kotanya sedikit (Syarifuddin et al., 2021). Dalam hal ini data jumlah kabupaten/kota di Provinsi Jawa Barat yaitu sebanyak 27.

3.3 Sumber Data

Sumber data yang digunakan adalah data sekunder, atau informasi yang dikumpulkan oleh orang lain dengan menggunakan berbagai alat atau teknik.(Hidayatullah et al., 2019). Data dalam penelitian ini diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Barat dalam buku data dan informasi kemiskinan provinsi tahun 2022 (Badan Pusat Statistik Provinsi Jawa Barat, 2021) yang dapat dilihat pada lampiran 1.

3.4 Variabel Penelitian

Variabel yang digunakan penulis dalam penelitian ini adalah 4 variabel yang didasarkan pada penelitian sebelumnya yaitu penelitian yang dilakukan Widyadhana et al., (2021) yang berjudul "Perbandingan Analisis Klaster *K-Means* dan *Average Linkage* untuk Pengklasteran Kemiskinan

di Provinsi Jawa Tengah”. Penulis mengambil variabel Garis kemiskinan, Jumlah Penduduk Miskin, Jumlah Penduduk Miskin Ekstrem, dan Pengeluaran Perkapita. Selain itu berdasarkan informasi dari BPS Jawa Barat ke empat variabel tersebut merupakan indikator kemiskinan. Dari variabel-variabel tersebut mempunyai pengertian masing-masing yaitu:

1. Yang dimaksud dengan penduduk miskin adalah seseorang yang rata-rata pendapatan bulannya kurang dari garis kemiskinan (TIM BPS Jawa Barat, 2023).
2. Garis kemiskinan adalah batas pengeluaran biaya kehidupan sehari-hari untuk kebutuhan sandang, pangan, pendidikan, dan kesehatan (TIM BPS Jawa Barat, 2023).
3. Penduduk miskin ekstrem adalah kondisi dimana masyarakat memiliki tingkat kemakmuran yang termasuk dalam garis kemiskinan ekstrem yaitu USD 1,9 PPP (*purchasing power parity*) (Akhmad Fatikhurrizqi, 2020).

4. Pengeluaran per kapita merupakan biaya konsumsi yang dikeluarkan oleh setiap anggota rumah tangga (Ningtyas et al., 2015).

3.5 Teknik Analisis Data

Dengan menggunakan RStudio dan Microsoft Excel untuk pengolahan data, metode *k-means* dan *average linkage* digunakan untuk menganalisis data penelitian. Langkah-langkah dalam proses analisis data adalah sebagai berikut:

- a) Mengumpulkan data penduduk miskin di Jawa Barat yang diperoleh dari BPS Provinsi Jawa Barat.
- b) Menentukan variabel dari data penduduk miskin di Jawa Barat.
- c) Melakukan analisis deskriptif.
- d) Melakukan standarisasi data.
- e) Melakukan pengujian analisis korelasi
- f) Melakukan uji *representative*
- g) Menghitung jarak *euclidean* setiap objek dengan objek lain.

- h) Melakukan proses pengelompokkan dengan metode *k-means*
- Menentukan banyak kluster k optimal dengan menggunakan metode *elbow*.
 - Menentukan ukuran suatu objek diposisikan di suatu kluster dengan menggunakan metode *silhouette*.
 - Menentukan jumlah kluster agar lebih konstan dengan menggunakan metode *gap statistic*.
 - Validasi penentuan kluster terbaik.
 - Menentukan *silhouette coefficient*.
 - Menentukan plot kluster kemiskinan di Jawa Barat.
 - Menentukan hasil pengklasteran kemiskinan di Jawa Barat.
- i) Melakukan proses pengelompokkan dengan metode *average linkage*
- Menentukan k optimal dengan metode *elbow*.

- Menentukan nilai k optimal pada metode *silhouette*.
 - Menentukan nilai k optimal dengan metode *gap statistic*.
 - Validasi penentuan kluster terbaik.
 - Menentukan *silhouette coefficient*.
 - Membuat kluster dendogram kemiskinan di Jawa Barat.
 - Menentukan hasil pengklasteran kemiskinan di Jawa Barat.
- j) Melakukan perbandingan metode *k-means* dan *average linkage* dengan menghitung nilai *silhouette coefficient*nya sehingga diperoleh metode yang terbaik.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisis Deskriptif

Analisis deskriptif digunakan dalam penelitian ini untuk memperoleh gambaran keseluruhan mengenai karakteristik kemiskinan di Provinsi Jawa Barat tahun 2022 berdasarkan variabel-variabel yang digunakan. Hasil analisis deskriptif yang ditampilkan pada **Tabel 4.1** adalah sebagai berikut. Untuk lebih lengkapnya bisa dilihat pada lampiran 2.

Tabel 4.1 Hasil Analisis Deskriptif

Variabel	<i>Mean</i>	<i>Min</i>	<i>Max</i>
Garis kemiskinan	Rp.468.619	Rp.335.134	Rp.744.771
Jumlah penduduk miskin	140.144 jiwa	9.400 jiwa	474.700 jiwa

Jumlah penduduk miskin ekstrem	33.624,4 jiwa	1.790 jiwa	146.120 jiwa
Pengeluaran perkapita	Rp.734.871	Rp.486.510	Rp.1.205.790

Pada **Tabel 4.1** terlihat bahwa garis kemiskinan tertinggi berada di Kota Depok sebesar Rp. 744.771. Selain itu, untuk jumlah penduduk miskin paling sedikit terjadi di Kota Banjar dengan penduduk miskin sebanyak 9.400 orang, sedangkan untuk jumlah penduduk miskin terbanyak terjadi di Kota Bogor dengan penduduk miskin sebanyak 474.700. Kemudian dari segi jumlah penduduk miskin ekstrem paling sedikit terjadi di Kota Banjar dengan penduduk miskin ekstrem sebanyak 1.790 orang, lalu untuk penduduk miskin ekstrem terbanyak terjadi di Kabupaten Bogor sebanyak 146.120 orang. Lalu untuk pengeluaran perkapita paling sedikit terjadi di Kabupaten Garut dengan pengeluaran perkapita sebanyak Rp. 486.510, sedangkan untuk pengeluaran perkapita terbanyak

terjadi di Kota Bandung dengan pengeluaran sebesar Rp. 1.205.790.

Jika dibandingkan dengan ke empat variabel kemiskinan di Provinsi Jawa Tengah lebih banyak angka kemiskinan di Jawa Barat pada tahun 2022. Pada variabel garis kemiskinan Jawa Barat yaitu untuk rata-ratanya sebesar 468.619, sedangkan untuk garis kemiskinan di Jawa Tengah yaitu mempunyai rata-rata sebesar 349.520. Untuk variabel jumlah penduduk miskin Jawa Barat juga lebih banyak yaitu sebanyak 4.070.000 jiwa (Badan Pusat Statistik Provinsi, 2023), sedangkan untuk jumlah penduduk miskin di Jawa Tengah sebanyak 3.831.440 jiwa. Untuk variabel jumlah penduduk miskin ekstrem di Jawa Barat adalah 941.860 jiwa (*BPS Jumlah Penduduk Miskin Ekstrem*), sedangkan untuk kasus penduduk miskin ekstrem di Jawa Tengah yaitu 639.123 (Anis Fauziah Titis, 2023). Kemudian untuk kasus pengeluaran perkapita per bulan di Jawa Barat yaitu sebanyak Rp. 734.871, sedangkan di Jawa Tengah sebanyak Rp. 771.120. Dari perbandingan

tersebut peneliti memilih Provinsi Jawa Barat dikarenakan tingkat kemiskinan di Jawa Barat tinggi jika dibandingkan dengan Provinsi lain di Indonesia.

4.2 Standarisasi Data

Apabila data yang digunakan dalam suatu penelitian mempunyai satuan yang berbeda, maka perlu dilakukan standarisasi. Selain itu, proses pengelompokan juga dipengaruhi oleh luasnya cakupan data yang sangat bervariasi. Karena data yang digunakan dalam penelitian ini berada dalam satuan yang berbeda, maka standarisasi data menjadi penting untuk memperkecil jangkauan data. Berikut adalah hasil standarisasi data 1 di variabel jumlah penduduk miskin (\bar{x}_1) dengan menggunakan bantuan persamaan (2.1) sampai (2.3).

a) Mencari \bar{x}_k

$$\begin{aligned}\bar{x} &= \frac{1}{N} \sum_{i=1}^N x_{ik} \\ \bar{x}_1 &= \frac{474.700 + 186.300 + \dots + 12.700}{27} \\ &= 140.144,44\end{aligned}$$

b) Mencari $x_{ik} - \bar{x}_i$

$$\begin{aligned} &= ((474700 - 140.144,44)^2 + (186300 - 140.144,44)^2 + \\ &\dots (12700 - 140.144,44)^2) / (27 - 1) \\ &= 5083,3607 \end{aligned}$$

c) Mencari σ_k

$$\begin{aligned} \sigma_k^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_{ik} - \bar{x}_i)^2 \\ &= \frac{(474700 - 140.144,44)^2 + (186300 - 140.144,44)^2 + \dots (12700 - 140.144,44)^2}{27 - 1} \\ &= \sqrt{249.912.305,0565} \end{aligned}$$

$$= 15.808,6149$$

d) Mencari \hat{x}_{ik}

$$\begin{aligned} \hat{x}_{ik} &= \frac{x_{ik} - \bar{x}_i}{\sigma_k} \\ &= \frac{15.808,6149}{5083,3607} \\ &= 3,12833695 \end{aligned}$$

Tabel 4.2 Hasil Standarisasi Data

	Jumlah.penduduk.miskin	Pengeluaran.perkapita	Jumlah.penduduk.miskin.ekstrem	Garis.kemiskinan
[1,]	3.128336950	-0.8327355	3.435630988	-0.2233268
[2,]	0.431587901	-0.8637358	-0.966120635	-0.9981151
[3,]	0.997306755	-0.8939105	0.077436244	-0.5557040
[4,]	1.107645308	-0.8383796	0.696485056	-0.6271563
[5,]	1.276893428	0.3299087	1.482588482	-1.2004840
[6,]	0.504523555	-0.8873398	-0.546193592	-1.0855217
[7,]	-1.222555327	-0.8597259	-0.972228665	-0.5695088
[8,]	0.001454557	0.7330804	-0.052970191	-0.8719470
[9,]	1.177775744	0.8092332	1.474037241	-0.4650058
[10,]	0.065039486	-0.8598776	-0.289961745	0.1321993
[11,]	-0.187430085	1.0210121	0.097592742	-0.8701034
[12,]	-1.100060831	-0.8469383	0.593259354	0.2804641
[13,]	0.141715430	1.0100609	0.055447337	-0.8661643
[14,]	-0.530601687	-0.8140848	-0.509850815	-0.5841770
[15,]	0.558757759	1.1232793	0.928284785	0.4724996
[16,]	0.569978629	-0.8157022	-0.238348893	0.9946819
[17,]	0.407276016	-0.8854528	-0.001051938	-0.6714757
[18,]	-0.956059668	0.7030911	-0.845792450	-0.5166277
[19,]	-0.569874732	1.3049003	-0.585590383	1.2620370
[20,]	-1.061722859	-0.8171680	-0.889770264	1.1286562
[21,]	-0.283742551	2.3496424	-0.252091960	0.6929902
[22,]	-1.015904307	1.2806392	-0.864421940	0.1528301
[23,]	-0.025662545	-0.7228531	0.858958647	2.3632128
[24,]	-0.708265460	-0.7220949	-0.490305120	2.4835351
[25,]	-1.018709525	-0.8013983	-0.946574940	0.7269402
[26,]	-0.496004005	0.8208583	-0.297596782	0.2706253
[27,]	-1.191697935	0.9756911	-0.950850561	-0.8553542

4.3 Uji Asumsi Korelasi

Salah satu asumsi yang harus dipenuhi ketika melakukan analisis kluster yaitu adalah tidak adanya korelasi antara tiap-tiap variabel (Wibowo et al., 2020). Data yang digunakan untuk analisis diperoleh melalui situs resmi BPS Jawa Barat. Terdapat beberapa variabel yang diamati, yaitu diantaranya jumlah penduduk miskin, pengeluaran perkapita, jumlah penduduk miskin ekstrem, dan garis

kemiskinan sebagai input dalam melakukan analisis kluster. Standarisasi data dilakukan sebelum pengelompokan untuk menyamakan satuan menjadi nilai standar dan tidak lagi mengandalkan satuan pengukuran. Analisis korelasi juga diuji untuk memastikan tidak ada hubungan yang signifikan antara dua kelompok variabel atau lebih. Berikut adalah perhitungan uji korelasi \bar{x}_1 dan \bar{x}_2 dengan bantuan persamaan (2.5).

a) Mencari \bar{x}_1

$$\begin{aligned}\bar{x}_1 &= 474.700 + 186.300 + \dots + 12.700 \\ &= 3.683.900\end{aligned}$$

b) Mencari \bar{x}_2

$$\begin{aligned}\bar{x}_1 &= 724.620 + 614820 + \dots + 716490 \\ &= 19.841.520\end{aligned}$$

c) Mencari nilai \bar{x}_1^2

$$\begin{aligned}\bar{x}_1^2 &= (474700)^2 + (186300)^2 + \dots + (12700)^2 \\ &= 13.751.119.210.000\end{aligned}$$

d) Mencari nilai \bar{x}_2^2

$$\bar{x}_2^2 = (72462)^2 + (614820)^2 + \dots + (716490)^2$$

$$= 525.074.144.400 + 378.003.632.400 + \dots + 5,1.335.792 \times 10^{11}$$

$$= 3,93.685.916 \times 10^{14}$$

e) Mencari nilai $r_{\bar{x}_1\bar{x}_2}$

$$r_{\bar{x}y} = \frac{(\sum(X-\bar{X})(Y-\bar{Y}))}{[\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2]^{\frac{1}{2}}}$$

$$r_{\bar{x}_1\bar{x}_2} =$$

$$\frac{(27.4,09.887.768 \times 10^{14}) - (3683900)(19841520)}{[(27.19841520 - (3683900))^2 (27.3,98.835.873 \times 10^{14} - (19841520))^2]^{\frac{1}{2}}}$$

$$= -0.27559104$$

Tabel 4.3 Hasil Uji Analisis Korelasi

	Penduduk Miskin (\bar{x}_1)	Pengeluaran Perkapita (\bar{x}_2)	Miskin ekstrem (\bar{x}_3)	Garis kemiskinan (\bar{x}_4)
Penduduk miskin (\bar{x}_1)	1	-0.2755910	0.79921824	-0.2751640

Pengeluaran perkapita (\bar{x}_2)	-0.275591	1	-0.0753661	0.78147335
Miskin ekstrim (\bar{x}_3)	0.7992182	-0.0753661	1	-0.0698587
Garis kemiskinan (\bar{x}_4)	-0.275164	0.78147335	-0.0698587	1

Jika nilai korelasi lebih dari 0,8 menandakan terjadinya hubungan yang terlalu kuat, maka perlu ditangani data tersebut (Widyadhana et al., 2021). Berdasarkan **Tabel 4.3** diperoleh bahwa:

1. Nilai korelasi pada \bar{x}_1 dan \bar{x}_2 dengan nilai 0,2755910 dinyatakan bahwa semakin tinggi penduduk miskin (\bar{x}_1) maka semakin rendah pengeluaran perkapita (\bar{x}_2). Hal ini menunjukkan bahwa \bar{x}_1 dan \bar{x}_2 tidak terjadi korelasi karena nilai korelasinya kurang dari 0,8.
2. Nilai korelasi pada \bar{x}_1 dan \bar{x}_3 adalah 0,7992182 yang artinya dinyatakan bahwa semakin tinggi penduduk miskin (\bar{x}_1) maka tinggi nilai

penduduk miskin ekstrem (\bar{x}_3). Hal ini menunjukkan bahwa tidak terjadinya korelasi karena nilai korelasinya kurang dari 0,8. Dengan demikian untuk uji asumsi korelasi terpenuhi. Selanjutnya akan dilanjutkan penelitian berikutnya.

4.4 Uji Asumsi *Representative*

Persamaan 2.6 digunakan untuk melakukan uji statistik *KMO (Kaiser-Mayer-Olkin)* berdasarkan anggapan tersebut guna menilai kelayakan data penelitian dari analisis yang akan dilakukan. Uji asumsi *representatif* menghasilkan hasil seperti di bawah ini, yang ditampilkan pada **Tabel 4.4**, dan untuk perhitungan R-Studio nya dapat dilihat pada lampiran 5.

Tabel 4.4 Hasil Uji *KMO*

<i>KMO</i> : 0,52
<i>KMO</i> pada tiap variabel
X_1 : 0,53
X_2 : 0,54
X_3 : 0,50
X_4 : 0,54

Berdasarkan **Tabel 4.4** diperoleh nilai *KMO* sebesar 0,52. Nilai tersebut berkisar antara 0,5 sampai 0,6, yang artinya sampel masuk dalam kategori baik atau dapat mewakili populasi. Hal ini dapat dinyatakan bahwa variabel-variabel tersebut layak digunakan untuk dianalisis lebih lanjut.

4.5 Perhitungan Jarak

Perhitungan jarak sebelumnya telah dilakukan untuk mengatasi masalah pengelompokan objek yang memiliki karakteristik serupa. Dengan mengacu pada data yang sudah terstandarisasi, maka

jarak yang dipilih penulis dalam penelitian ini adalah jarak *euclidean*.. Berikut merupakan contoh perhitungan jarak antara Kabupaten Bogor dan Kabupaten Sukabumi (objek 1 dan 2) menggunakan persamaan (2.5). Berdasarkan hasil perhitungan di bawah diperoleh jarak *euclidean* antara Kabupaten Bogor dan Kabupaten Sukabumi (objek 1 dan 2) menggunakan persamaan (2.4).

$$d_{1,2} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_4 - y_4)^2} =$$

$$\sqrt{(3,12833695 - 0,431587901)^2 + (-0,8327355 + 0,8637358)^2} +$$

$$\sqrt{(3,435630988 + 0,96612063) + (-0,223326815 + 0,998115051)^2}$$

$$= \sqrt{27,6284941559} = 5,2562814$$

Perhitungan jarak antara objek 1 dan objek 2 menghasilkan nilai jarak *euclidean* sebesar 5,2562814. Untuk perhitungan jarak antara Kabupaten Bogor dengan Kabupaten Cianjur (objek 1 dan 3) sebagai berikut.

$$d_{1,3} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_4 - y_4)^2}$$

$$= \sqrt{(3,12833695 - 0,997306755)^2 + (-0,8327355 + 0,8939105)^2}$$

$$+$$

$$\sqrt{(3,435630988 - 0,077436244)^2 + (-0,223326815 + 0,555703962)^2}$$

$$= \sqrt{17,426521276} = 4,1745085071$$

Perhitungan jarak *euclidean* antara objek 1 dan objek 3 didapatkan nilai jarak *euclidean* sebesar 4,1745085071. Hasil penghitungan jarak euclidean menggunakan R-Studio 4.1.3 ditunjukkan di bawah ini. Lampiran 5 berisi seluruh hasil perhitungan jarak *euclidean*.

Tabel 4.4 Hasil Perhitungan Jarak *Euclidean*

	1	2	3	4	...	27
1	0	5,2562814	4,1745085	3,4295683	...	6,1891409
2	5,2562814	0	1,4046858	1,9007227	...	1,7267770
3	4,1745085	1,4046858	0	1,2783776	...	2,7067556
4	3,4295683	1,9007227	1,2783776	0	...	2,8385363
5	3,1600031	2,6964854	1,5752139	1,5731038	...	3,7152653
...
27	6,1891409	1,7267770	2,7067556	2,8385363	...	0

Berdasarkan hasil perhitungan di atas diperoleh jarak *euclidean* antara Kabupaten Bogor dengan Kabupaten Cianjur memiliki karakteristik lebih mirip dibandingkan dengan Kabupaten Sukabumi. Demikian pula, semakin dekat dua objek satu sama lain saat menghitung objek lain, semakin mirip sifat-sifatnya. Langkah selanjutnya adalah melakukan analisis klaster menggunakan metode *kmeans* dan *average linkage*, dengan bantuan aplikasi R-Studio 4.1.3 digunakan untuk mengolah data tersebut.

4.6 Analisis *K-means* Klaster

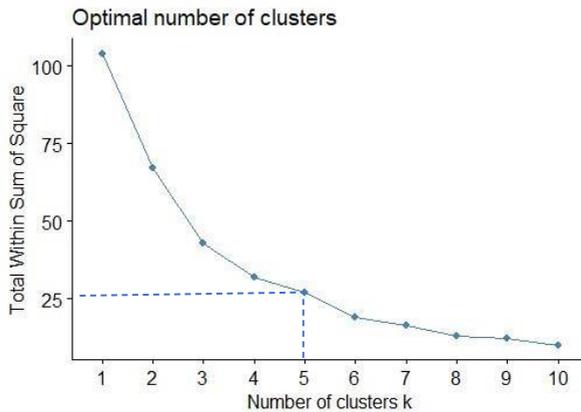
Prosedur analisis metode *k-means* melibatkan pengelompokan objek dengan properti serupa, dan dimulai dengan menghitung jumlah klaster.

Kemudian melakukan analisis berikut:

- Metode *Elbow*

Pencarian nilai klaster yang ideal dilakukan setelah menentukan jarak ke setiap objek. Pendekatan siku digunakan dalam metode *k-means* untuk mengidentifikasi

jumlah kluster yang ideal, khususnya dengan memfokuskan pada suatu titik yang membentuk siku. Hasil analisis metode *elbow* dari 1 kluster menjadi 10 kluster adalah sebagai berikut.



Gambar 4.1 Visualisasi k optimal metode *Elbow*

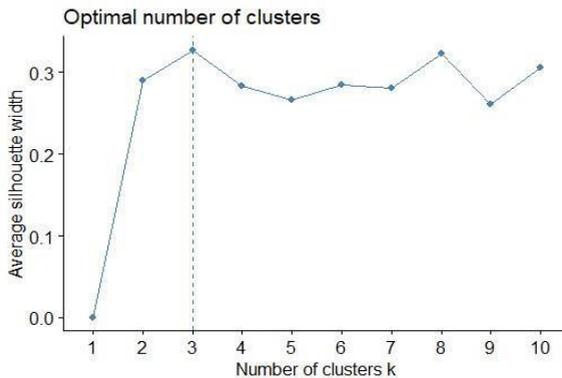
Titik bergaris berbentuk siku pada **Gambar 4.1** ditampilkan, dengan titik lima memiliki nilai kluster terbaik. Menurut metode *elbow*, lima kluster adalah jumlah kluster yang ideal karena, setelah poin kelima, penurunannya stabil atau tidak ada penurunan yang terlihat landai. Untuk mendapatkan hasil

yang akurat, nilai ini harus dihitung ulang menggunakan uji kinerja kluster.

- Metode *silhouette*

Uji performa pada metode *k-means* kluster dilakukan dengan menggunakan metode *silhouette*. Berikut ini ditampilkan grafik visualisasi kluster optimal dengan menggunakan metode *silhouette* pada

Gambar 4.2.

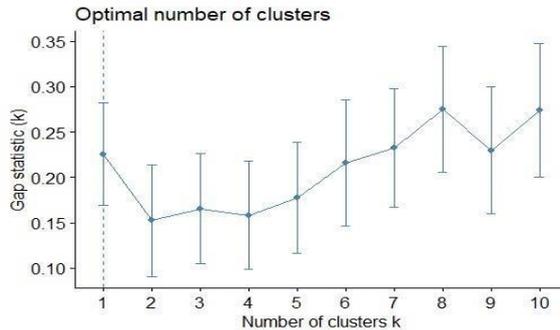


Gambar 4.2 Visualisasi k optimal metode *silhoutte*

Berdasarkan **Gambar 4.2** menunjukkan bahwa kluster 3 menempati titik tertinggi, sehingga menurut metode *silhouette* maka

diperoleh kluster yang optimal sebanyak 3 kluster.

- Metode *gap statistic*

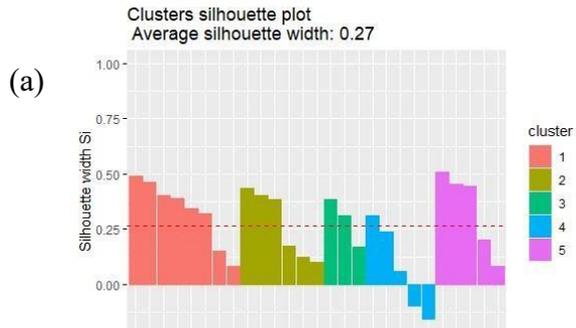


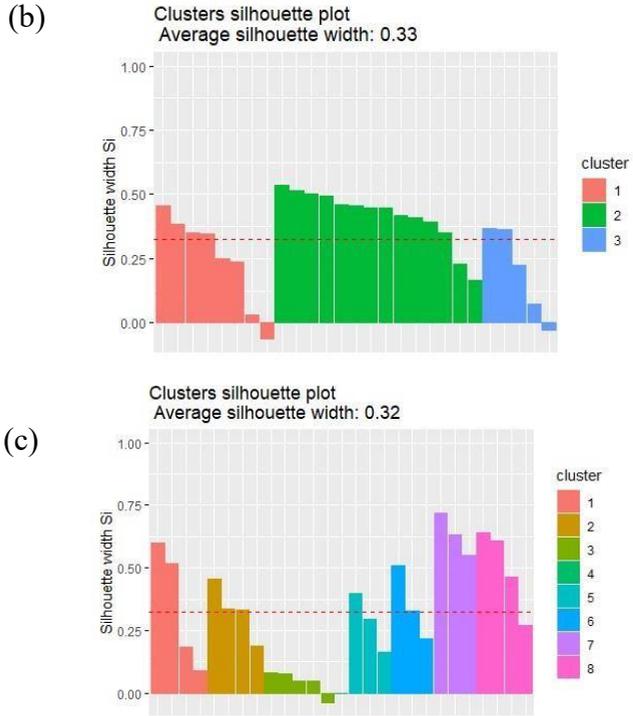
Gambar 4.3 Visualisasi K optimal Metode Gap statistic

Berdasarkan **Gambar 4.3** menunjukkan bahwa kluster 8 dan 10 menempati titik tertinggi, sehingga boleh memilih antara 8 kluster atau 10 kluster, tetapi disini penulis memilih 8 kluster untuk kluster optimalnya dalam metode *gap statistic*, sehingga menurut metode *gap statistic* maka diperoleh kluster yang optimal sebanyak 8 kluster.

- Plot *Cluster Silhouette Plot Average Silhouette Width* (Membandingkan Jumlah Kluster Optimal masing-masing Metode)

Setelah dilakukan penelitian ketiga metode tersebut untuk menentukan kluster optimalnya, di dapatkan masing-masing metode berbeda yaitu dengan metode *elbow* adalah 5 kluster, kemudian dengan metode *silhouette* 3 kluster, dan menggunakan metode *gap statistic* 8 kluster. Selanjutnya adalah melakukan tahap validasi bahwa hasil kluster mewakili populasi secara umum.

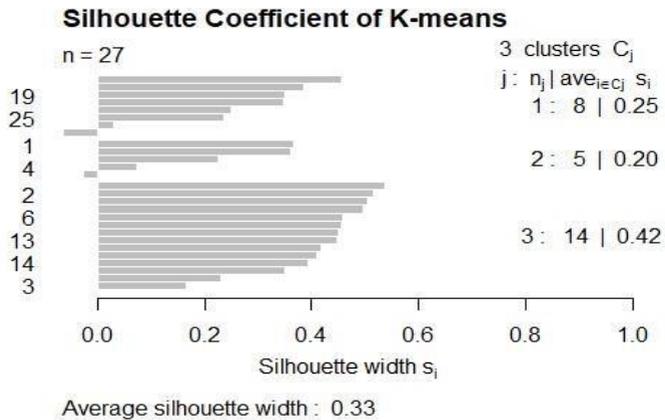




Gambar 4.4 Validasi Nilai Ke tiga Metode

Berdasarkan **Gambar 4.4** Nilai *Average Silhouette Width* dari metode *elbow* (a) adalah 5 klaster, metode *silhouette* (b) 3 klaster, dan dengan metode *gap statistic* (c) 8 klaster, jumlah klaster optimal yang tepat adalah 3 klaster, karena setelah di validasi

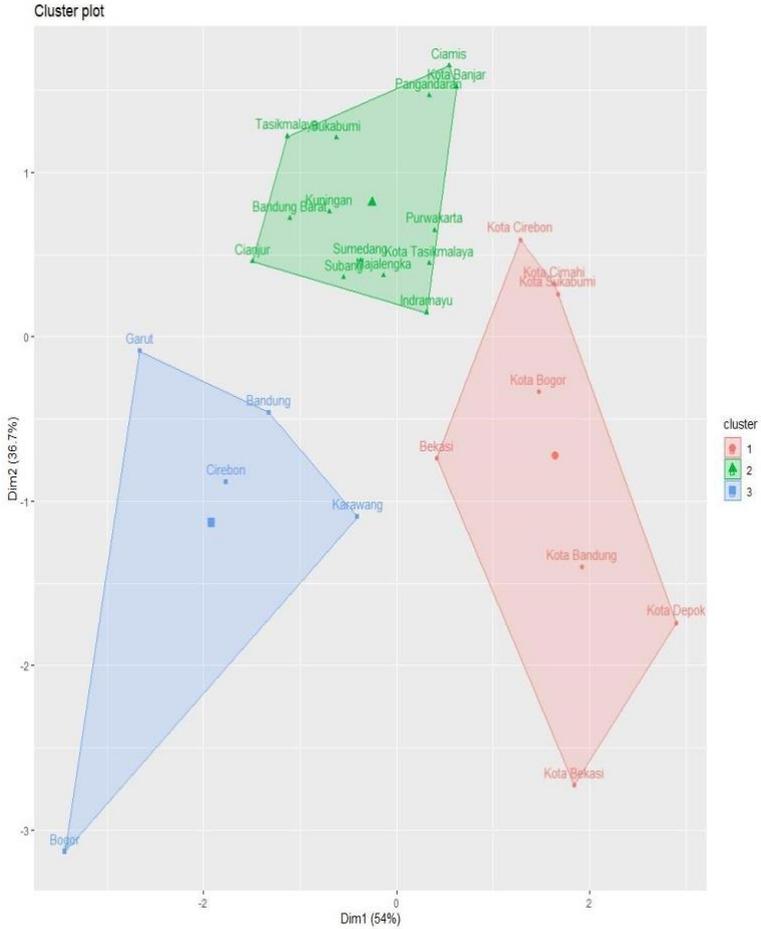
metode *silhouette* menunjukkan nilai *average silhouette width*nya yang paling besar yaitu 0,33 yang artinya semua objek berada dalam kluster yang tepat. Langkah berikutnya adalah menghitung nilai *silhouette coefficient* yang terbentuk dengan kluster sama dengan 3.



Gambar 4.5 Hasil Uji Nilai *Silhouette Coefficient* Metode *K-means*

Berdasarkan **Gambar 4.5** menunjukkan Mayoritas wilayah Jawa Barat berada pada kluster sesuai dengan nilai rata-rata *silhouette coefficient* sebesar 0,33 yang menunjukkan bahwa kluster secara

keseluruhan baik. Di Provinsi Jawa Barat, kluster kemiskinan dibentuk setelah diperoleh kluster ideal. Berikut adalah plot yang dihasilkan menggunakan analisis kluster *k-means*.



Gambar 4.6 Kluster Plot Kemiskinan

Gambar 4.6 menunjukkan bahwa pengelompokan klaster berdasarkan 27 Kabupaten/Kota di Jawa Barat terbentuk ke dalam 3 klaster yang terbentuk dalam klaster warna merah adalah kelompok di klaster pertama, kemudian kelompok di warna hijau adalah klaster ke dua, dan kelompok warna biru termasuk ke dalam klaster 3. Sehingga hasil pengklasteran di Jawa Barat terdapat pada **tabel 4.5**.

Tabel 4.5 Hasil Pengklasteran Kemiskinan Di Jawa Barat Dengan Metode *K-means*

Klaster	Kabupaten/Kota
Klaster 1	Bekasi, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi
Klaster 2	Sukabumi, Cianjur, Tasikmalaya, Ciamis, Kuningan, Majalengka, Sumedang, Indramayu, Subang,

	Purwakarta, Bandung Barat, Pangandaran, Kota Tasikmalaya, Kota Banjar
Klaster 3	Bogor, Bandung, Garut, Cirebon, Karawang

Berdasarkan **Tabel 4.5** setelah diidentifikasi sesuai klaster dan datanya. Untuk Kabupaten/Kota yang berada di dalam Klaster 1: mempunyai anggota yang memiliki pengeluaran perkapita yang tinggi di atas rata-rata dan Garis kemiskinan yang tinggi di atas rata-rata.

Klaster 2: memiliki rata-rata yang rendah dibanding dengan klaster 3 dan lebih tinggi dibanding klaster 1 untuk setiap variabelnya.

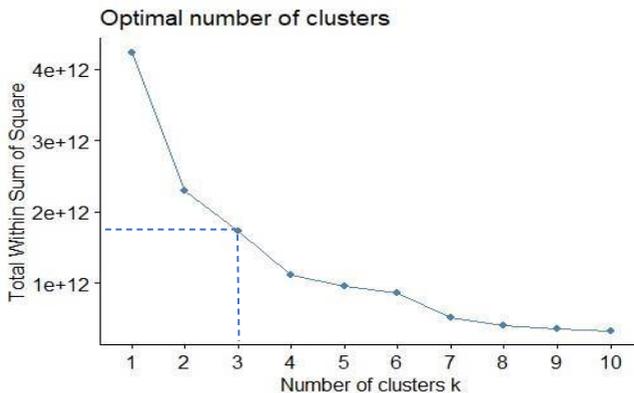
Klaster 3: memiliki jumlah penduduk miskin dan jumlah penduduk miskin ekstrem yang tinggi di atas rata-rata.

4.7 Metode *Average Linkage*

Pertama, dengan menggunakan tiga teknik yang sama seperti metode *k-means* sebelumnya, yaitu menentukan jumlah kluster yang ideal yaitu sebagai berikut:

- Metode *elbow*

Dalam menentukan nilai kluster yang optimal pada metode *elbow* dapat dilihat dari grafik yang penurunannya mulai landai.



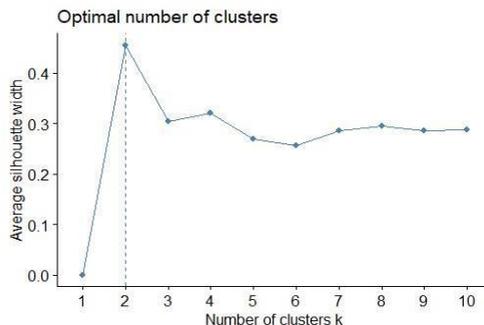
Gambar 4.7 Visualisasi kluster optimal metode *elbow*

Ditampilkan titik bergaris berbentuk siku pada **Gambar 4.7**. Titik ketiga merupakan titik yang

mempunyai nilai kluster terbaik, karena Menurut metode *elbow*, tiga kluster merupakan jumlah kluster yang ideal karena titik tersebut berbentuk siku, selain dari itu untuk setelah poin ketiga, penurunan titik mulai landai. Maka dari itu kluster ketiga merupakan kluster yang terbentuk pada penelitian metode *elbow*. Untuk mendapatkan hasil yang akurat, nilai ini harus dihitung ulang menggunakan uji kinerja kluster.

- Metode *silhouette*

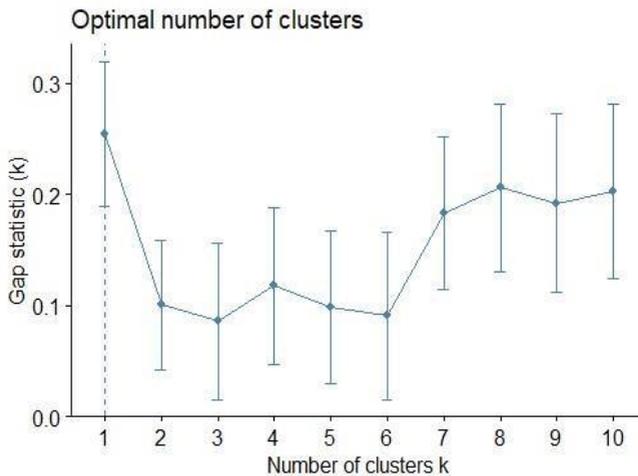
Garis tertinggi atau garis paling ideal dapat digunakan untuk menentukan nilai kluster ideal bila menggunakan pendekatan *silhouette*.



Gambar 4.8 Visualisasi kluster optimal metode *silhouette*

Berdasarkan **Gambar 4.8** menunjukkan bahwa kluster 2 menempati titik tertinggi, sehingga menurut metode *silhouette* maka diperoleh kluster yang optimal sebanyak 2 kluster.

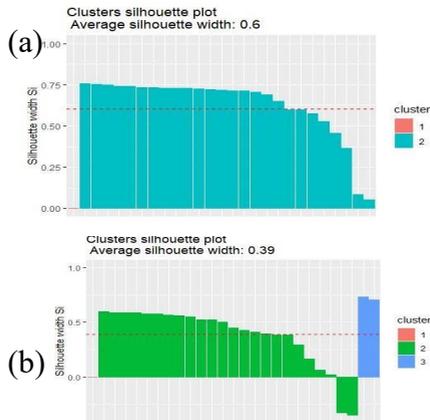
- Metode *gap statistic*
Penentuan nilai kluster optimal dapat dilihat dari garis yang tertinggi.



Gambar 4.9 Visualisasi kluster optimal metode *gap statistic*

Berdasarkan **Gambar 4.9** menunjukkan bahwa kluster 1 menempati titik tertinggi, sehingga 1

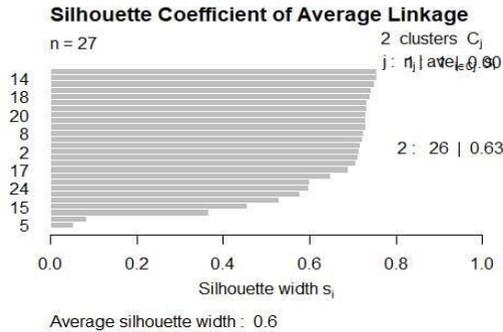
kluster merupakan kluster optimal dalam metode *gap statistic*, maka diperoleh kluster yang optimal adalah sebanyak 1 kluster. Setelah dilakukan nilai kluster optimal dengan metode untuk menentukan jumlah kluster optimalnya, diperoleh kluster optimal yang diperoleh dengan metode *elbow* sebanyak 3 kluster, metode *silhouette* sebanyak 2 kluster, dan metode *gap statistic* sebanyak 1 kluster. Proses validasi adalah langkah berikutnya untuk memastikan hasil kluster mencerminkan populasi yang lebih luas secara akurat.



Gambar 4.10 Hasil uji pemilihan kluster terbaik metode *average linkage*

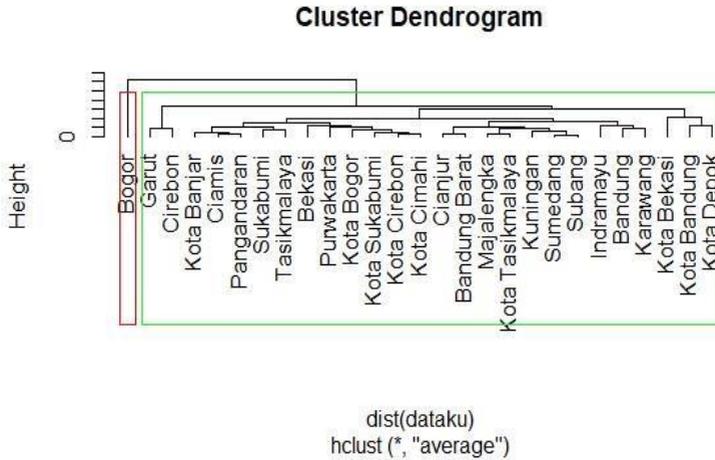
Berdasarkan Hasil **Gambar 4.10** kluster dalam metode gap statistik sama dengan 1 tidak penulis masukkan ke dalam validasi grafiknya, dikarenakan kluster 1 belum bisa dikatakan kluster, melainkan populasi. Maka nilai *Average Silhouette Width* dari metode *elbow* adalah 3 kluster dan metode *silhouette* 2 kluster, dan jumlah kluster optimal yang tepat adalah 2, karena setelah di validasi metode *silhouette* menunjukkan nilai *average silhouette width* nya yang paling besar yaitu 0,6 Hal ini menunjukkan bahwa setiap objek berada dalam klasternya masing-masing. Langkah terakhir adalah menghitung *silhoutte coefficient* yaitu kluster sama dengan 2.

- Plot *Silhouette Coefficient* of *Average Linkage*



Gambar 4.11 Hasil uji *silhouette coefficient* dengan metode *average linkage*

Berdasarkan hasil **Gambar 4.11** di atas menunjukkan rata-rata nilai *silhouette coefficient* nya adalah 0,6 hal ini menunjukkan semua wilayah di Jawa Barat masuk dalam klaster yang tepat, dalam hal ini *silhoutte coefficient* yang terbentuk adalah sebesar 0,6. Setelah memperoleh klaster yang optimal, maka langkah selanjutnya adalah menentukan pengklasteran dengan *dendogram*.



Gambar 4.12 Kluster *dendogram* kemiskinan di Jawa Barat

Berdasarkan **Gambar 4.12** dengan identifikasi *dendogram* yang terbentuk diatas dengan metode *average linkage* dapat disimpulkan bahwa wilayah yang berada di kotak warna merah merupakan kluster 1, dan wilayah yang berada di kotak berwarna hijau merupakan kluster 2.

Tabel 4.6 Hasil Pengklasteran Di Jawa Barat
Dengan Metode *Average Linkage*

Klaster	Kabupaten\Kota
Klaster 1	Bogor
Klaster 2	Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Kota Bogor, Kota, Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.

Setelah menerapkan metode *average linkage* untuk menganalisis klaster kemiskinan di Jawa Barat didapatkan hasil di **Tabel 4.6** dengan klaster 1

diidentifikasi sebagai tingkat jumlah penduduk miskin dan penduduk miskin ekstrem yang paling tinggi dibandingkan klaster 2, sedangkan untuk kabupaten/kota yang berada dalam klaster 2 yaitu sebaliknya yang berada dalam klaster 1.

4.8 Perbandingan Hasil Klaster *K-means* dan *Average Linkage*

Diperoleh tiga klaster dengan pendekatan *kmeans* dan dua klaster dibuat menggunakan metode *average linkage* berdasarkan proses analisis klaster yang telah dilakukan. Langkah selanjutnya adalah membandingkan pendekatan *average linkage* dan *k-means* klaster untuk menemukan teknik analisis klaster yang paling efektif. Penentuan metode terbaik untuk pengelompokan kemiskinan di Provinsi Jawa Barat berdasarkan data kemiskinan BPS Provinsi Jawa Barat pada tahun 2022 dilakukan dengan menghitung nilai *silhouette coefficient*. Berikut ini adalah data perbandingan untuk metode *k-means* dan *average linkage*.

Tabel 4.7 Hasil Perbandingan Metode *K-means* dan *Average Linkage*

Metode <i>K-means</i>	Metode <i>Average Linkage</i>
Nilai <i>silhouette coefficient</i> nya adalah 0,33	Nilai <i>silhouette coefficient</i> nya adalah 0,6

Pada **Tabel 4.7** menunjukkan bahwa nilai *silhouette coefficient* pada metode *average linkage* lebih besar daripada metode *k-means* yaitu sebesar 0,6. Dengan demikian metode *average linkage* adalah metode terbaik untuk pengelompokkan kemiskinan di Provinsi Jawa Barat pada tahun 2022 berdasarkan penjelasan *sillhoutte coefficient* dan penjelasan pada **Tabel 2.2**.

BAB V

PENUTUP

5.1 Kesimpulan

Kesimpulan berikut dapat diambil berdasarkan temuan dan pembahasan pada bab sebelumnya, yaitu sebagai berikut:

- a. Berdasarkan hasil pengelompokan 27 Kabupaten/Kota di Provinsi Jawa Barat pada data kemiskinan pada tahun 2022 menggunakan metode *k-means* menghasilkan kluster yang terbentuk secara optimal sebanyak 3 kluster, yaitu pada kluster 1 mempunyai anggota sebanyak 8 kabupaten/kota, kluster 2 mempunyai anggota sebanyak 14 kabupaten/kota, dan kluster 3 mempunyai anggota sebanyak 5 kabupaten/kota. Untuk hasil pengelompokan dengan metode *average linkage* diperoleh kluster yang optimal sebanyak 2 kluster, yaitu kluster 1 mempunyai jumlah anggota

sebanyak 27 kabupaten/kota, dan klaster 2 mempunyai jumlah anggota 1 kabupaten/kota.

b. Gambaran karakteristik pada 3 klaster menggunakan metode *k-means* klaster sebagai berikut:

- Klaster 1 merupakan klaster dengan kabupaten/kota yang mempunyai pengeluaran perkapita dan garis kemiskinan di atas rata-rata.
- Klaster 2 merupakan klaster dengan kabupaten/kota yang mempunyai jumlah penduduk miskin, jumlah penduduk miskin ekstrem, garis kemiskinan, dan pengeluaran perkapitanya itu ada yang lebih tinggi dan ada pula yang lebih rendah dari rata-rata per variabelvariabel tersebut.
- Klaster 3 merupakan klaster dengan kabupaten/kota yang mempunyai jumlah penduduk miskin dan jumlah

penduduk miskin ekstrem diatas rata-rata, namun untuk garis kemiskinannya berada dibawah rata-rata.

Gambaran karakteristik pada 2 klaster menggunakan metode *average linkage* klaster sebagai berikut:

- Klaster 1 merupakan klaster dengan kabupaten/kota yang mempunyai jumlah penduduk miskin dan jumlah penduduk miskin ekstrem yang tinggi diatas rata-rata, tetapi untuk variabel garis kemiskinan dan pengeluaran perkapita rendah dibawah rata-rata.
- Klaster 2 merupakan klaster dengan kabupaten/kota yang mempunyai jumlah penduduk miskin, jumlah penduduk miskin ekstrem, pengeluaran perkapita, dan garis kemiskinan yang merata yaitu

ada yang diatas rata-rata dan ada yang dibawah rata-rata.

- c. Metode terbaik dari perbandingan hasil pengelompokkan kemiskinan di Provinsi Jawa Barat pada tahun 2022 antara metode *kmeans* dan *average linkage* adalah *average linkage* karena metode tersebut menghasilkan nilai *silhouette coefficient* terbesar.

5.2 Saran

Berdasarkan temuan-temuan tersebut, diharapkan para peneliti berikutnya dapat memperluas penelitian ini dengan menggunakan tambahan teknik analisis kluster terhadap informasi yang secara signifikan mempengaruhi pertumbuhan kemiskinan di suatu wilayah, dan pemerintah diharapkan dapat segera menyelesaikan permasalahan kemiskinan yang difokuskan kepada kabupaten/kota yang teridentifikasi.

DAFTAR PUSTAKA

- Adelia, O, S., & Rita, H, B. 2019. Analisis Korelasi Sektor Pertanian Terhadap Tingkat Kemiskinan di Provinsi Sumatera Utara. Sumatera Utara: Jurnal Agribisnis Sumatera Utara.
- Agoestanto, A., Sukestiyarno, Y. L., & Rochmad. (2016). *Analysis of Mathematics Critical Thinking Students in Junior High School based on Cognitive Style. Journal Physics: Conference Series*, 824.
- Anindita, A. 2017. Dampak Pertumbuhan Ekonomi, Jumlah Pengangguran Terhadap Tingkat Kemiskinan Di Kabupaten Sidoarjo. In Seminar Nasional; FEB Unikama “Peningkatan Ketahanan Ekonomi Nasional Dalam Rangka Menghadapi Persaingan Global (pp. 130-137).
- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, 4(1), 51. <https://doi.org/10.30865/mib.v4i1.1784>.
- Agama, K., & Karawang, K. (2010). Jurnal Manajemen Vol.10 No.1 Oktober 2012 1038. *Jurnal Manajemen*, 10(1), 1038–1046.

- Akhmad Fatikhurizqi, B. D. K. (2020). *Peran Bantuan Sosial dalam Pengentasan Kemiskinan Ekstrem di Jawa Timur Tahun 2020 (The Role of Social Assistance in Alleviating Extreme Poverty in East Java in 2020)*. 2020, 1027–1036.
- Anis Fauziyah Titis. (2023). *639.123 Penduduk Jateng Miskin Ekstrem, 4629 Anak Tidak Bersekolah*. Kompas.Com. <https://www.google.com/amp/s/amp.kompas.com/regional/read/2023/02/14/050000278/639-123penduduk-jateng-miskin-ekstrem4-629-anak-tidakbersekolah>.
- Badan Pusat Statistik Indonesia. (2023). Profil kemiskinan di indonesia september 2023. *Berita Resmi Statistik*, 01(05), 1–16.
- Badan Pusat Statistik Provinsi. (2023). Jumlah Penduduk Miskin. In *Badan Pusat Statistik* (pp. 2012–2023). <https://jatim.bps.go.id/subject/162/produkdomeestikregional-bruto-kabupatenkota.html#subjekViewTab5>.
- Badan Pusat Statistik Provinsi Jawa Barat. (2021). *Jumlah Penduduk Miskin*. Badan Pusat Statistik.
- Bangun, R. H. B. (2016). Analisis Klaster Non Heirarki Dalam Pengelompokan Kabupaten/Kota di Sumatera Utara Berdasarkan Faktor Produksi Padi. *Jurnal Agribisnis Sumatera Utara*, 4(1), 54–61.

- bps.go.id. (2023). *Badan Pusat Statistik*. Bps.Go.Id.[https://www.bps.go.id/subject/5/konsumsidanpengeluaran.html#:~:text=Pengeluaran rata-rata per kapita,tangga dalam rumah tangga tersebut](https://www.bps.go.id/subject/5/konsumsidanpengeluaran.html#:~:text=Pengeluaran%20rata-rata%20per%20kapita,tangga%20dalam%20rumah%20tangga%20tersebut)
- bps.go.id. (2023). *Konsep Kemiskinan*. Bps.Go.Id.<https://www.bps.go.id/subject/23/kemiskinandanketimpangan.html>
- BPS Jawa Barat. (2022). Profil Kemiskinan di Jawa Barat Maret 2022. *Website BPS Jawa Barat*, 43.
- Cahya, B. T. (2015). Kemiskinan Ditinjau Dari Perpekstif Al-Quran Dan Hadis. *Jurnal Penelitian*, 9(1), 41–66. <https://doi.org/10.21043/jupe.v9i1.850>
- Cebeci, Z., & Yildiz, F. (2015). Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures. *Journal of Agricultural Informatics*, 6(3), 13–23. <https://doi.org/10.17700/jai.2015.6.3.196>
- Dewa, F. A., & Jatipaningrum, M. T. (2019). SEGMENTASI E-COMMERCE DENGAN CLUSTER K-MEANS DAN FUZZY C-MEANS (Studi Kasus: Media Sosial di Indonesia yang diunduh di Play Store). *Jurnal Statistika Industri Dan Komputasi*, 4(1), 53–67.
- Dewi, S. W. K., Utami, B. O., & Musafa. (2021). Mengukur Tingkat Kepuasan Pengguna ECommerce Shopee Pada

Fitur Cod (Cash on Delivery). *Jurnal Sains Manajemen*, 3(2), 55–63. <https://doi.org/10.51977/jsm.v3i2.561>

Dewi, V. E. (2021). Apakah Pertumbuhan Ekonomi Selaras dengan Penurunan Kemiskinan (Studi Kasus di Provinsi Jambi). *Jurnal Ilmiah Populer Apakah*, 4,16–23. <https://bpsjambi.id/median/index.php/median/article/download/11/9>

Ediyanto, Mara, M. N., & Satyahadewi, N. (2013). Pengklasifikasian Karakteristik Dengan Metode KMeans Cluster Analysis. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, 02(2), 133–136.

Farisi, A., Herlambang, A. D., & Zulvarina, P. (2023). Hubungan Minat Belajar dan Motivasi Belajar Siswa terhadap Perencanaan Karier Siswa Jurusan Teknik Komputer dan Jaringan di SMK Negeri 3 Malang. 7(4), 1872–1882.

Goejantoro, R., & Desi Yuniarti, dan. (2019). Perbandingan Pengelompokan K-Means dan KMedoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus : Data Titik Panas Di Indonesia Pada 28 April 2018). *Jurnal EKSPONENSIAL*, 10(2), 143–152.

Goreti, M., Novia N, Y., & Wahyuningsih, S. (2016). Perbandingan hasil analisis cluster dengan menggunakan metode single linkage dan metode cmeans (studi kasus: data tingkat kualitas udara ambien

pada perusahaan perkebunan di Kabupaten Kutai Barat tahun 2014). *Jurnal Eksponensial*, 7(1), 9–16.

- Hadyan Tisantri, D., Cahya Wihandika, R., & Adinugroho, S. (2019). Prediksi Keputusan Pelanggan Menggunakan Extreme Learning Machine Pada Data Telco Customer Churn. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer**Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(11), 10516– 10523.
- Halilintar R, & Farina Nur I. (2018). Penerapan Algoritma K-Means Clustering Untuk Prediksi Prestasi Nilai Akademik Mahasiswa. *Jurnal Sains Dan Informatika*, 4(2).
- Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering. *Jurnal Ilmiah Teknologi Dan Rekayasa*, 25(1), 76–88. <https://doi.org/10.35760/tr.2020.v25i1.2677>.
- Handoyo, R., Mangkudjaja, R., & Nasution, S. M. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K - Means pada Pengelompokan Dokumen. *Jurnal SIFO Mikroskil*, 15(2), 73–82. <https://doi.org/10.55601/jsm.v15i2.161>.
- Harahap, B. (2019). Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung). *Regional Development Industry & Health Science, Technology*

and Art of Life, 394–403.
<https://ptki.ac.id/jurnal/index.php/readystar/article/view/82>

- Hidayat. (2017). Penanggulangan Kemiskinan di Jawa Barat. *Jurnal Ilmiah Magister Ilmu Administrasi (JIMIA)*, 2, 66–72. *Jumlah Penduduk Miskin Ekstrem (1)*. (n.d.).
- Lestari, S. P., Supandi, E. D., & Rahayu, P. P. (2018). Pengklasteran Kabupaten/Kota di Jawa Tengah berdasarkan Tenaga Kesehatan dengan Menggunakan Metode Ward dan K-Means. *Jurnal Fourier*, 7(2), 103–109. <https://doi.org/10.14421/fourier.2018.72.103-109>.
- Mulia, R. A. (2022). Pengaruh Tingkat Kemiskinan dan Produk Domestik Regional Bruto Terhadap Kesejahteraan Masyarakat. *Jiee: Jurnal Ilmiah Ekotrans & Erudisi*, 2(1), 22–27.
- Mulya, Di. P. (2019). Analisa Dan Implementasi Association Rule Dengan Algoritma Fp-Growth Dalam Seleksi Pembelian Tanah Liat (Studi Kasus Di Pt. Anveve Ismi Berjaya). *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 1(1), 47–57. <https://doi.org/10.47233/jteksis.v1i1.6>.
- Muningsih, E., & Kiswati, S. (2018). Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan. *Joutica*, 3(1), 117. <https://doi.org/10.30736/jti.v3i1.196>.
- Nafisah, Q., & Chandra, N. E. (2017). Analisis Cluster Average Linkage Berdasarkan Faktor-Faktor Kemiskinan di Provinsi Jawa Timur. *Zeta – Math*

Journal, 3(2), 31-

36.<https://doi.org/10.31102/zeta.2017.3.2.31-36>.

Nasmiwati, R., & Triani, M. (2019). Analisis Faktor Faktor Yang Mempengaruhi Kemiskinan Rumah Tangga Di Kecamatan Bayang Kabupaten Pesisir Selatan. *Jurnal Kajian Ekonomi Dan Pembangunan*, 1(1), 213. <https://doi.org/10.24036/jkep.v1i1.5647>.

Ningrat, D. R., Asih, D., Maruddani, I., & Wuryandari, T.(2016). Analisis Cluster Dengan Algoritma K-Means Dan Fuzzy C-Means Clustering Untuk Pengelompokan Data Obligasi Korporasi. *Jurnal Gaussian*, 5(4), 641–650.<http://ejournals1.undip.ac.id/index.php/gaussian>.

Ningtyas, R., Rahmawati, R., & Wilandari, Y. (2015). Penerapan Metode Empirical Best Linear Unbiased Prediction (EBLUP) pada Model Penduga Area Kecil dalam Pendugaan Pengeluaran per Kapita di Kabupaten Brebes. *Jurnal Gaussian*, 4(4), 977–986. <http://ejournals1.undip.ac.id/index.php/gaussian>.

Nugroho, M. R., Hendrawan, I. E., & Purwantoro, P. P. (2022). Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI. *Nuansa Informatika*, 16(1), 125–133. <https://doi.org/10.25134/nuansa.v16i1.5294>.

Palupi, R., Yulianna, D. A., & Winarsih, S. S. (2021). Analisa Perbandingan Rumus Haversine Dan Rumus Euclidean Berbasis Sistem Informasi Geografis Menggunakan

Metode Independent Sample t-Test. *JITU: Journal Informatic Technology And Communication*, 5(1), 40–47. <https://doi.org/10.36596/jitu.v5i1.494>.

Paramadina, M., Sudarmin, S., & Aidid, M. K. (2019). Perbandingan Analisis Cluster Metode Average Linkage dan Metode Ward (Kasus: IPM Provinsi Sulawesi Selatan). *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 1(2), 22. <https://doi.org/10.35580/variansium9357>.

Prahartiwi, L. I., & Dari, W. (2019). Algoritma Apriori untuk Pencarian Frequent itemset dalam Association Rule Mining. *PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic*, 7(2), 143–152. <https://doi.org/10.33558/piksel.v7i2.1817>.

Pramesthy, T. D., Hutapea, R. Y. F., & Tesen, M. (2021). Pramesthy. dkk 24 ANALISIS PENGARUH LAMA SETTING DAN LAMA PENARIKAN TALI KERUT TERHADAP TOTAL HASIL TANGKAPAN PURSE SEINE DI SIBOLGA ANALYSIS OF THE EFFECT SETTING TIME AND TIME TO PULL OF PURSE LINE FOR THE TOTAL CATCH OF PURSE SEINE IN SIBOLGA. *Jurnal IPTEKS PSP*, 8(1), 24–33.

Pusporini, W., Widodo, S. A., Wijayanti, A., Wijayanti, N., Utami, W. B., Taqiyuddin, M., & Irfan, M. (2023). Mathematical Knowledge Content in Junior High School Curriculum: A Comparative Study of the 2013

Curriculum and Merdeka Curriculum. *Mosharafa: Jurnal Pendidikan Matematika*, 12(2), 389–404.
<https://doi.org/10.31980/mosharafa.v12i2.2210>

Putriana, U., Setyawan, Y., & Noeryanti. (2016). Metode Cluster Analysis Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Variabel Yang Mempengaruhi Kemiskinan Pada Tahun 2013. *Jurnal Statistika Industri Dan Komputasi*, 1(1), 38–52.

Qadrini, L. (2020). Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 13(2), 5–11.
<https://doi.org/10.36456/jstat.vol13.no2.a2886>

Sihombing, A. O., & Bangun, R. H. (2019). Analisis Korelasi Sektor Pertanian Terhadap Tingkat Kemiskinan di Provinsi Sumatera Utara. *Jurnal Agrica*, 12(1), 17.
<https://doi.org/10.31289/agrica.v12i1.2220>.

Suraya, G. R., & Wijayanto, A. W. (2022). Comparison of Hierarchical Clustering, K-Means, K-Medoids, and Fuzzy C-Means Methods in Grouping Provinces in Indonesia according to the Special Index for Handling Stunting. *Indonesian Journal of Statistics and Its*

Applications, 6(2), 180–201.

<https://doi.org/10.29244/ijsa.v6i2p180-201>.

Suyanto, Syarippudin, & Wasono. (2021). Di Kabuapten Kutai Kartanegara Tahun 2019 Single Linkage Cluster Analysis Based on Village Potential. *Eksponensial*, 12(1), 59–64.

Syarifuddin, Bata Ilyas, Jamaluddin, & Sani, A. (2021). Pengaruh Persepsi Pendidikan dan Pelatihan Sumber Daya Manusia Pada Kantor Dinas Di Kota Makassar. *Bata Ilyas Educational Management Review*, 1(2), 51–56. <https://ojs.stieamkop.ac.id/index.php/biemr/article/view/102>.

TIM BPS Jawa Barat. (2023). *BPS Provinsi Jawa Barat*. <https://jabar.bps.go.id/subject/23/kemiskinan.html#subjekViewTab1>

Wahyuni, R., Nugroho, S., Novianti, D. P., Statistika, M. J., Matematika, P., Pembimbing, D., & Statistika, J. (2013). Analisis Klaster Dengan Menggunakan Metode Single Linkage Dan Metode K-Means (Studi Kasus: Data Produksi, Produktivitas, Dan Luas Panen Tanaman Sayuran Di Provinsi Bengkulu Tahun 2013). *Prodi Matematika, FMIPA, UNIB, Bengkulu*, 1–9.

[http://sigitnugroho.id/eSkripsi/2015/12/Analisis Klaster Metode Single Linkage dan k-Means.pdf](http://sigitnugroho.id/eSkripsi/2015/12/Analisis%20Klaster%20Metode%20Single%20Linkage%20dan%20k-Means.pdf).

Wardy, D. K., Putra, I. K. G. D., & Rusjyanthi, N. K. D. (2022). Clustering Artikel pada Portal Berita Online. *JITTER- Jurnal Ilmiah Teknologi Dan Komputer*, 3(1), 3–11.

Wibowo, R., Nisa, K., Faisol, A., & Setiawan, E. (2020). Simulasi Pemilihan Metode Analisis Cluster Hirarki Agglomerative Terbaik Antara Average Linkage Dan Ward Pada Data Yang Mengandung Masalah Multikolinearitas. *Jurnal Siger Matematika*, 1(2), 49–55. <https://doi.org/10.23960/jsm.v1i2.2497>.

Widayati, T. (2013). Analisis Tingkat Kemiskinan Di Kabupaten Demak. *MEDIA EKONOMI DAN MANAJEMEN*, 28(2), 44–54.

Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah. *PRISMA, Prosiding Seminar Nasional Matematika*, 4, 584–594. <https://journal.unnes.ac.id/sju/index.php/prisma/>

Wijaya, T., & Budiman, S. (2016). *Analisis Multivariat Untuk Penelitian Manajemen*. 21-2-2018 Wulan Permata Sari, & Tata Sutabri. (2023). Analisa Cluster Dengan K-Mean Clustering Untuk Pengelompokan Data

Cybercrime. *Jurnal Informatika Teknologi Dan Sains*, 5(1), 49–53. <https://doi.org/10.51401/jinteks.v5i1.2209>

Yacoub, Y., & Mutiaradina, H. (2020). Analisis kesejahteraan petani dan kemiskinan perdesaan di Indonesia. *COVID-19 Pandemic, Mitigate The Shock and Pave The Way For A Sustainable Future, 2017*, 1–11.

Yuhelmi, Mariza Devega, R. (2018). UJI LINEARITAS MENGGUNAKAN STATISTICAL PRODUCT AND SERVICE SOLUTIONS (SPSS) UNTUK MAHASISWA SEMESTER VII (TUJUH) FAKULTAS ILMU BUDAYA UNIVERSITAS LANCANG KUNING. *Transcommunication*, 53(1), 1–9. <http://www.tfd.org.tw/opencms/english/about/background.html><http://dx.doi.org/10.1016/j.cirp.2016.06.001><http://dx.doi.org/10.1016/j.powtec.2016.12.055><https://doi.org/10.1016/j.ijfatigue.2019.02.006><https://doi.org/10.1016/j.matlet.2019.04.024>

Yusfar, A. A., Tiro, M. A., & Sudarmin, S. (2021). Analisis Cluster Ensemble dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Kinerja Pembangunan Ekonomi Daerah. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 3(1), 13. <https://doi.org/10.35580/variansiunm14626>

Zulyanti, T., & Noeryanti. (2022). Perbandingan pengelompokan usaha mikro kecil dan menengah di kabupaten klaten tahun 2019 dengan metode kmeans

dan clustering large application. *Jurnal Statistika Industri Dan Komputasi*, 7(1), 46–59.

LAMPIRAN-LAMPIRAN

Lampiran 1 Data Kemiskinan Jawa Barat 2022

Kabupaten /Kota	Jumlah penduduk miskin (ribu)	Pengeluaran perkapita	Jumlah penduduk miskin ekstrem (ribu)	Garis kemiskinan
Bogor	474,7 jiwa	Rp.724.620	146,12 jiwa	Rp.443.787
Sukabumi	186,3 jiwa	Rp.614.820	1,99 jiwa	Rp.357.636
Cianjur	246,8 jiwa	Rp.506.760	36,16 jiwa	Rp.406.829
Bandung	258,6 jiwa	Rp.704.520	56,43 jiwa	Rp.398.884
Garut	276,7 jiwa	Rp.486.510	82,17 jiwa	Rp.335.134
Tasikmalaya	194,1 jiwa	Rp.530.160	15,74 jiwa	Rp.347.917
Ciamis	9,4 jiwa	Rp.628.500	1,79 jiwa	Rp.405.294
Kuningan	140,3 jiwa	Rp.630.090	31,89 jiwa	Rp.371.665
Cirebon	266,1 jiwa	Rp.657.210	81,89 jiwa	Rp.416.914
Majalengka	147,1 jiwa	Rp.627.960	24,13 jiwa	Rp.483.319
Sumedang	120,1 jiwa	Rp.732.630	36,82 jiwa	Rp.371.870
Indramayu	22,5 jiwa	Rp.674.040	53,05 jiwa	Rp.499.805
Subang	155,3 jiwa	Rp.728.730	35,44 jiwa	Rp.372.308

Purwakarta	83,4 jiwa	Rp.791.040	16,93 jiwa	Rp.403.663
Karawang	199,9 jiwa	Rp.769.050	64,02 jiwa	Rp.521.158
Bekasi	201,1 jiwa	Rp.785.280	25,82 jiwa	Rp.579.221
Bandung Barat	183,7 jiwa	Rp.536.880	33,59 jiwa	Rp.393.956
Pangandaran	37,9 jiwa	Rp.619.410	5,93 jiwa	Rp.411.174
Kota Bogor	79,2 jiwa	Rp.833.730	14,45 jiwa	Rp.608.949
Kota Sukabumi	26,6 jiwa	Rp.780.060	4,49 jiwa	Rp.594.118
Kota Bandung	109,8 jiwa	Rp.1.205.790	25,37 jiwa	Rp.545.675
Kota Cirebon	31,5 jiwa	Rp.25.090	5,32 jiwa	Rp.485.613
Kota Bekasi	137,4 jiwa	Rp.1.115.940	61,75 jiwa	Rp.731.392
Kota Depok	64,4 jiwa	Rp.1.118.640	17,57 jiwa	Rp.744.771
Kota Cimahi	31,2 jiwa	Rp.836.220	2,63 jiwa	Rp.549.450
Kota Tasikmalaya	87,1 jiwa	Rp.661.350	23,88 jiwa	Rp.498.711
Kota Banjar	12,7 jiwa	Rp.716.490	2,49 jiwa	Rp.373.510

Lampiran 2 Output Statistik Deskriptif

```
> summary(data)
Kabupaten.Kota Jumlah.penduduk.miskin Pengeluaran.perkapita Jumlah.penduduk.miskin.ekstrem
Length:27      Min. : 9400      Min. : 486510    Min. : 17900
Class :character 1st Qu.: 51150    1st Qu.: 628230    1st Qu.: 101900
Mode :character  Median :137400   Median : 716490    Median : 253700
                Mean :140144    Mean : 734871     Mean : 336244
                3rd Qu.:197000  3rd Qu.: 788160    3rd Qu.: 449350
                Max. :474700     Max. :1205790     Max. :1461200

Garis.kemiskinan
Min. :335134
1st Qu.:383733
Median :416914
Mean :468619
3rd Qu.:533417
Max. :744771
.
```

Lampiran 3 Output Standarisasi Data

```

> #STANDARISASI DATA
> dataku=scale(dataku)
> dataku

```

	Jumlah.penduduk.miskin	Pengeluaran.perkapita	Jumlah.penduduk.miskin.ekstrem	Garis.kemiskinan
[1,]	3.128336950	-0.05757656	3.435630988	-0.2233268
[2,]	0.431587901	-0.67428102	-0.966120635	-0.9981151
[3,]	0.997306755	-1.28121256	0.077436244	-0.5557040
[4,]	1.107645308	-0.17047054	0.696485056	-0.6271563
[5,]	1.276893428	-1.39494904	1.482588482	-1.2004840
[6,]	0.504523555	-1.14978374	-0.546193592	-1.0855217
[7,]	-1.222555327	-0.59744571	-0.972228665	-0.5695088
[8,]	0.001454557	-0.58851529	-0.052970191	-0.8719470
[9,]	1.177775744	-0.43619266	1.474037241	-0.4650058
[10,]	0.065039486	-0.60047868	-0.289961745	0.1321993
[11,]	-0.187430085	-0.01258746	0.097592742	-0.8701034
[12,]	-1.100060831	-0.34166500	0.593259354	0.2804641
[13,]	0.141715430	-0.03449226	0.055447337	-0.8661643
[14,]	-0.530601687	0.31547909	-0.509850815	-0.5841770
[15,]	0.558757759	0.19196970	0.928284785	0.4724996
[16,]	0.569978629	0.28312738	-0.238348893	0.9946819
[17,]	0.407276016	-1.11204008	-0.001051938	-0.6714757
[18,]	-0.956059668	-0.64850075	-0.845792450	-0.5166277
[19,]	-0.569874732	0.55525244	-0.585590383	1.2620370
[20,]	-1.061722859	0.25380865	-0.889770264	1.1286562
[21,]	-0.283742551	2.64497066	-0.252091960	0.6929902
[22,]	-1.015904307	0.50672487	-0.864421940	0.1528301
[23,]	-0.025662545	2.14031770	0.858958647	2.3632128
[24,]	-0.708265460	2.15548256	-0.490305120	2.4835351
[25,]	-1.018709525	0.56923781	-0.946574940	0.7269402
[26,]	-0.496004005	-0.41293986	-0.297596782	0.2706253
[27,]	-1.191697935	-0.10323965	-0.950850561	-0.8553542

Lampiran 4 Output Uji Korelasi

```
1 #uji korelasi#
2 library(stats)
3 #Masukkan data#
4 JPM <-c(474700, 186300, 246800, 258600,276700,194100,9400,140300,266100,147100,12010
5 PP <-c(724620,614820,506760,704520,486510,530160,628500,630090,657210,627960,732630,
6 PME <-c(146120,1990,36160,56430,82170,15740,1790,31890,81890,24130,36820,53050,35440
7 GK <-c(443787,357636,406829,398884,335134,347917,405294,371665,416914,483319,371870,
8 #lakukan uji variabel 1 dengan variabel lainnya#
9 data_korelasi <-data.frame(JPM,PP,PME<GK)
10 with(data_korelasi,cor(JPM,PP))
11 with(data_korelasi,cor(JPM,PME))
12 with(data_korelasi,cor(JPM,GK))
13 with(data_korelasi,cor(JPM,JPM))
14 with(data_korelasi,cor(PP,PME))
15 with(data_korelasi,cor(PP,GK))
16 with(data_korelasi,cor(PME,GK))
17 #hasil keseluruhan uji korelasi
18 cor(data_korelasi, method = "pearson")
19 |
```

```
R 4.3.1 .- /
> library(stats)
> JPM <-c(4747.00, 1863.00, 2468.00, 2586.00, 2767.00, 1941.00, 94.00, 1403.00, 2661.00,
1471.00, 1201.00, 225.00, 1553.00, 834.00, 1999.00, 2011.00, 1837.00, 379.00, 792.00, 26
6.00, 1098.00, 315.00, 1374.00, 644.00, 312.00, 871.00, 127.00)
> PP <-c(724620, 614820, 506760, 704520, 486510, 530160, 628500, 630090, 657210, 627960,
732630, 674040, 728730, 791040, 769050, 785280, 536880, 619410, 833730, 780060, 1205790,
825090, 1115940, 1118640, 836220, 661350, 716490)
> JPME <-c(14612.0, 199.0, 3616.0, 5643.0, 8217.0, 1574.0, 179.0, 3189.0, 8189.0, 2413.0,
3682.0, 5305.0, 3544.0, 1693.0, 6402.0, 2582.0, 3359.0, 593.0, 1445.0, 449.0, 2537.0, 53
2.0, 6175.0, 1757.0, 263.0, 2388.0, 249.0)
> GK <-c(443787, 357636, 406829, 398884, 335134, 347917, 405294, 371665, 416914, 483319,
371870, 499805, 372308, 403663, 521158, 579221, 393956, 411174, 608949, 594118, 545675, 4
85613, 731392, 744771, 549450, 498711, 373510)
> data_korelasi <-data.frame(JPM,PP,JPME<GK)
> with(data_korelasi,cor(JPM,PP))
[1] -0.275591
> with(data_korelasi,cor(JPM,JPME))
[1] 0.7992182
> with(data_korelasi,cor(JPM,GK))
[1] -0.275164
> with(data_korelasi,cor(JPM,JPM))
[1] 1
> with(data_korelasi,cor(PP,JPME))
[1] -0.07536614
> with(data_korelasi,cor(PP,GK))
[1] 0.7814733
> with(data_korelasi,cor(JPME,GK))
[1] -0.06985877
```

Lampiran 5 Output Uji *Representative*

```
> library(psych)
> KMO(zdata)
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = zdata)
Overall MSA = 0.52
MSA for each item =
      Jumlah.penduduk.miskin      Pengeluaran.perkapita
              0.53              0.54
Jumlah.penduduk.miskin.ekstrem      Garis.kemiskinan
              0.50              0.54
.
```

Lampiran 6 Output Jarak Euclidean

```

> data=read.delim("clipboard")
> jarak=dist(x=dataku2,method="euclidean")
> dataku2=scale(x=dataku[,2:5])
> jarak

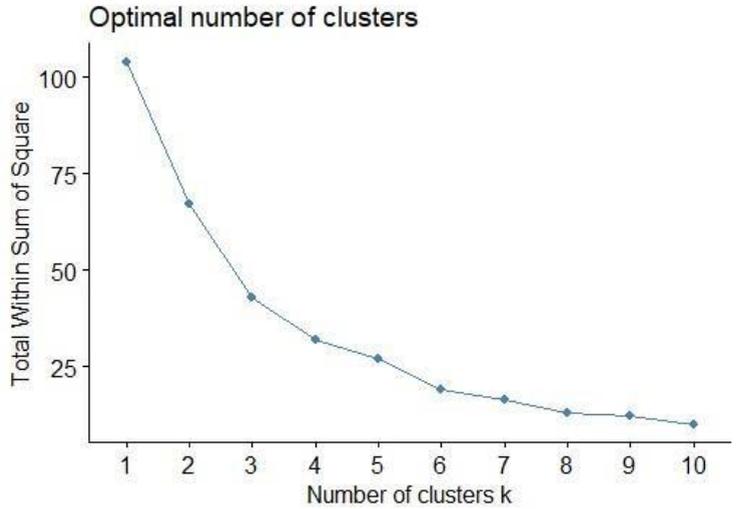
```

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	5.2562814														
3	4.1745085	1.4046858													
4	3.4295683	1.9007227	1.2783776												
5	3.1600031	2.6964854	1.5752139	1.5731038											
6	4.9674361	0.6445162	0.9642550	1.7541833	2.1876550										
7	6.2266195	1.7105070	2.5489825	2.8982787	3.6479463	1.9327840									
8	4.7592369	1.0208489	1.2603835	1.4212758	2.1778330	0.9257344	1.5603797								
9	8.2025220	2.6176437	1.6447943	0.8404801	1.2124518	2.3299900	3.4325981	1.9759268							
10	4.8667211	1.3693625	1.3930948	1.6797714	2.6491040	1.4294889	1.6173410	1.0337606	2.1756650						
11	4.7494373	1.4031731	1.7641611	1.4556979	2.4662789	1.4942828	1.6273962	0.6245341	2.0253226	1.2506662					
12	5.1276638	2.5540187	2.4993901	2.3953505	3.1244471	2.5282438	1.8037836	1.7377960	2.5551849	1.4921479	1.5845118				
13	4.3562399	1.2467429	1.5437672	1.1914549	2.2995594	1.3362613	1.8227105	0.5817237	1.8461066	1.2009400	0.3325783	1.8001004			
14	5.4059266	1.5116087	2.2868418	2.0921505	3.2468834	1.8631014	1.2354108	1.3787632	2.7264388	1.3248823	0.8222749	1.6494176	0.9866960		
15	3.6655109	2.5530456	2.0356171	1.3021577	2.4779339	2.5307628	2.9144443	1.9209741	1.3980461	1.5721517	1.7582013	1.7848151	1.6670559	2.0944589	
16	4.6521984	2.3316677	2.2656534	1.9997776	3.3310731	2.5454929	2.6408065	2.1451455	2.4392198	1.3350144	2.0618914	2.0931022	1.9578959	1.9439219	1.2814617
17	4.5307497	1.1091776	0.6295388	1.3658689	1.8213486	0.6924559	1.9684357	0.6940126	1.8080259	1.0527028	1.2695098	2.0309942	1.1281523	1.7843350	1.9737615
18	5.9538506	1.4739336	2.2516311	2.6226375	3.3811657	1.6727187	0.3039886	1.2943147	3.1593257	1.3322439	1.4177792	1.6796160	1.5863560	1.1080123	2.6696040
19	5.6964340	2.7870720	3.0939373	2.9246395	4.1898940	3.0941984	2.2931973	2.5440506	3.3558866	1.7614653	2.3412534	1.8543814	2.4071413	1.8636727	2.0786950
20	6.1798476	2.7604914	3.2200079	3.2380943	4.3864004	3.0730812	1.9081574	2.5578318	3.6903659	1.8308733	2.4094149	1.8096589	2.5306375	1.8341290	2.5230234
21	5.7779742	3.8598968	4.3271039	3.5362769	5.0349674	4.2744736	3.6752053	3.6090784	3.9937551	3.3121762	3.1044214	3.2359179	1.1442136	2.8804973	2.8581333
22	6.0104177	2.1965942	0.9391932	2.8307006	4.0265621	2.5866012	1.3398855	1.9857472	3.3987407	1.6506875	1.7131166	1.6935900	1.8755022	0.9704083	4.2738833
23	5.3016865	4.7707948	4.6780538	3.9487763	5.2233717	4.9974080	4.5696831	4.3295655	4.0575622	3.7120704	3.9617143	3.4238879	3.9789848	3.7610830	2.7779866
24	6.5082428	4.6335058	4.9273882	4.4488720	5.8318716	5.0136745	4.1709034	4.4140400	4.7777051	3.7097494	4.0699278	3.5237599	4.1275645	3.5816779	3.3937283
25	6.1398193	2.5740822	1.9096559	3.0986734	4.3293904	2.9530657	1.7561724	2.3950149	3.6215376	1.8241727	2.1611063	1.8457499	2.2919656	1.4874135	2.4921022
26	5.2386171	1.7278172	1.9512412	2.1034954	3.0725770	1.8560521	1.3125825	1.2820317	2.5450614	0.6075859	1.3087900	1.0787420	1.4024405	1.1434727	1.7383834
27	6.1891409	1.7267770	2.7067556	2.8385363	3.7152653	2.0467404	0.5721507	1.5702133	3.4289526	1.7996309	1.4547242	1.9338052	1.6719658	0.9382918	2.9061296
2		16	17	18	19	20	21	22	23	24	25	26			
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															
13															
14															
15															
16															
17	2.1921121														
18	2.4186295	1.6766387													
19	1.2511487	2.7955043	2.1976155												
20	1.7622736	2.8379210	1.8799522	0.6656566											
21	1.5294961	4.0641634	3.6214113	2.2099388	2.6304653										
22	1.9146024	2.4638851	1.3366557	1.2285700	1.0094281	2.4030973									
23	2.6231479	4.5512876	4.4545263	2.4714184	3.0355428	2.0845520	3.3919075								
24	2.7239287	4.7026162	4.1292945	2.0201548	2.3951127	1.9190141	2.8957128	1.5169598							
25	1.7829916	2.7766366	1.7445430	0.7863121	0.5157017	2.3091779	0.5833243	3.0646001	2.4302996						
26	1.4658093	1.5100162	1.0896908	1.4172828	1.3606902	3.0945653	1.2046767	3.5294317	3.4023455	1.3664882					
27	2.6801261	2.1237525	0.6918169	2.3317453	2.0209911	3.3559912	1.1945148	4.4752398	4.0860515	1.7279539	1.5081362				

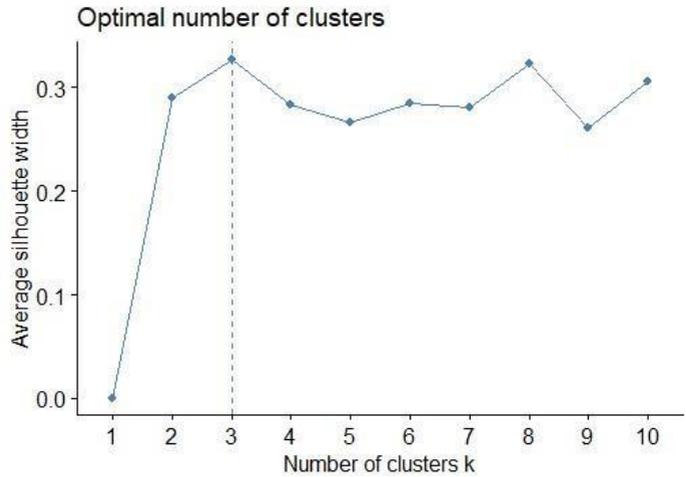
Lampiran 7 Output Analisis *K-Means*

```
1 library(cluster)
2 library(factoextra)
3 library(tidyverse)
4
5 set.seed(135)
6 data=read.delim("clipboard")
7 dataku=data[,-1]
8 #Merubah data menjadi numerik
9 str(dataku)
10 dataku$Jumlah.penduduk.miskin=as.numeric(dataku$Jumlah.penduduk.miskin)
11 dataku$Pengeluaran.perkapita=as.numeric(dataku$Pengeluaran.perkapita)
12 dataku$Jumlah.penduduk.miskin.ekstrem=as.numeric(dataku$Jumlah.penduduk.miskin.ekstrem)
13 dataku$Garis.kemiskinan=as.numeric(dataku$Garis.kemiskinan)
14 str(dataku)
15
16 #STANDARISASI DATA
17 dataku=scale(dataku)
18 str(dataku)
19 #Melihat data dan merubah data menjadi numerik
20 view(dataku)
21 summary(dataku)
22 row.names(dataku)=data[,1]
23 #K-Optimal Cluster
24 fviz_nbclust(dataku, kmeans, method = "silhouette")
25 fviz_nbclust(dataku, kmeans, method = "wss")
26 fviz_nbclust(dataku, kmeans, method = "gap_stat")
27 #Pembuatan model k means
28 final=kmeans(dataku, 5)
29 final
30 # Menghitung silhouette coefficient
31 silhouette_result=silhouette(final$cluster, dist(dataku))
32 silhouette_result
33 #plot silhouette cluster #BUAT 3 BENTUK
34 fviz_silhouette(silhouette_result)
35 # Menampilkan nilai silhouette coefficient
36 print(mean(silhouette_result[, "sil_width"]))
37 #Plot cluster
38 fviz_cluster(final, data = dataku)
39
40 # Menghasilkan objek silhouette
41 silhouette_obj <- silhouette(final$cluster, dist(dataku))
42 # Plot koefisien silhouette
43 plot(silhouette_obj, main = "silhouette coefficient of k-means")
44 ..
```

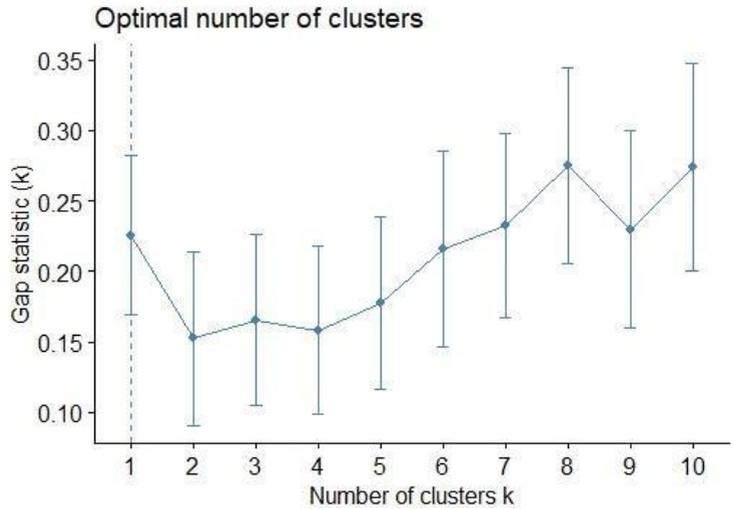
- Plot Metode Elbow



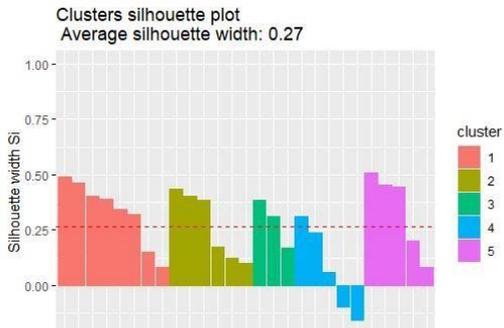
- Plot Metode Silhouette



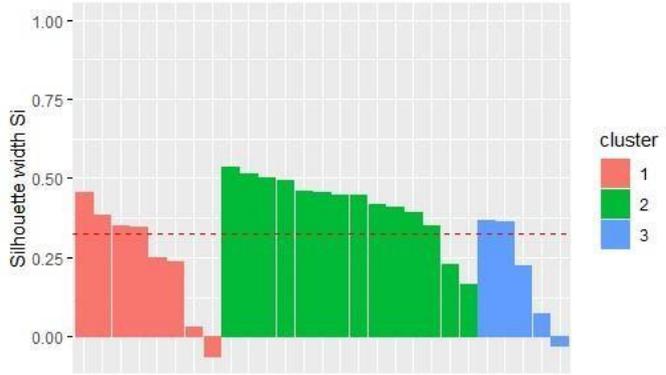
Plot Metode Gap Statistics



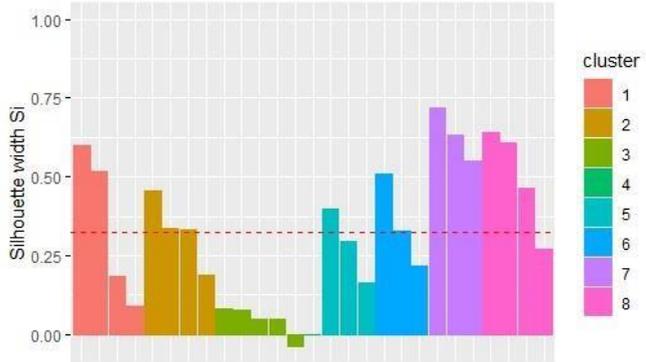
- Plot Cluster Silhouette Plot Average Silhouette Width.



Clusters silhouette plot
Average silhouette width: 0.33

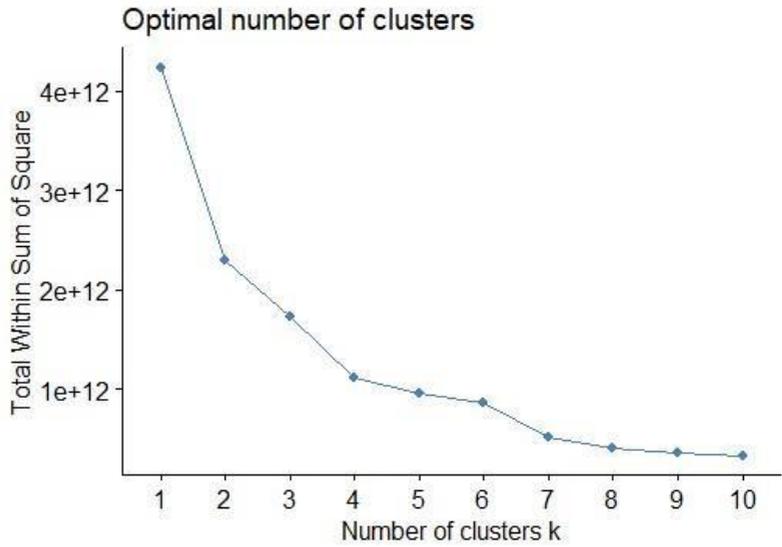


Clusters silhouette plot
Average silhouette width: 0.32

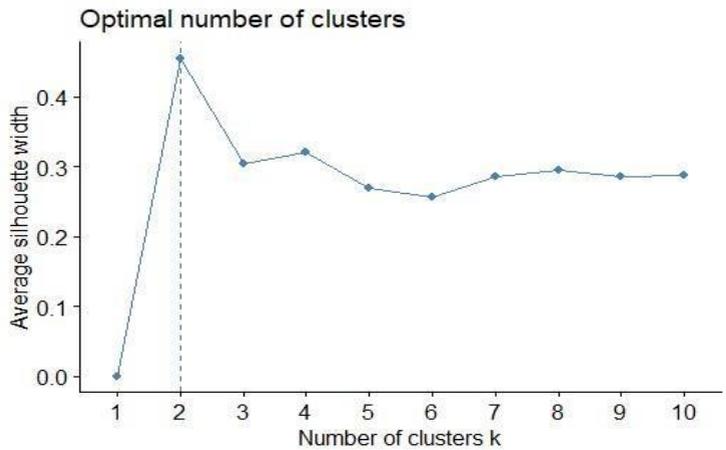


Lampiran 8 Analisis *Average Linkage* Plot Elbow Average Linkage

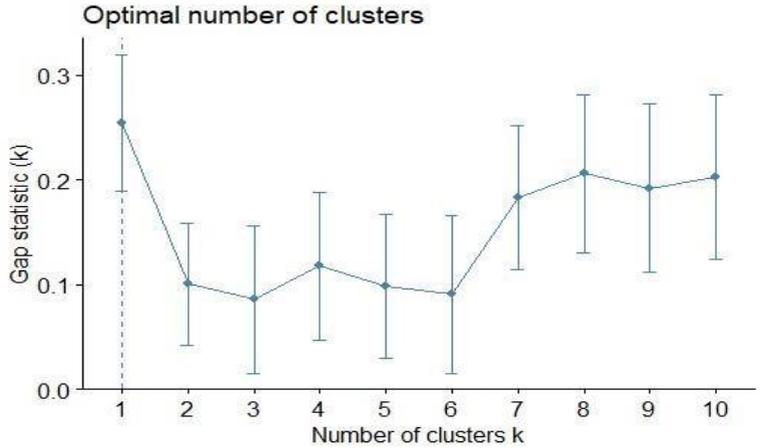
```
1 # Instal dan muat paket yang diperlukan
2 library(cluster)
3
4 # Persiapkan data
5 set.seed(135)
6 data=read.delim("clipboard")
7 dataku=data[,-1]
8 row.names(dataku)=data[,1]
9 view(dataku)
10 #Merubah data menjadi numerik
11 str(dataku)
12 dataku$Jumlah.penduduk.miskin=as.numeric(dataku$Jumlah.penduduk.miskin)
13 dataku$Pengeluaran.perkapita=as.numeric(dataku$Pengeluaran.perkapita)
14 dataku$Jumlah.penduduk.miskin.ekstrem=as.numeric(dataku$Jumlah.penduduk.miskin.ekstrem)
15 dataku$Garis.kemiskinan=as.numeric(dataku$Garis.kemiskinan)
16 str(dataku)
17
18 # Lakukan pengelompokan hierarkis menggunakan metode average linkage
19 #K-Optimal Cluster
20 fviz_nbclust(dataku, kmeans, method = "silhouette")
21 fviz_nbclust(dataku, kmeans, method = "wss")
22 fviz_nbclust(dataku, kmeans, method = "gap_stat")
23
24 hc <- hclust(dist(dataku), method = "average")
25 # Hitung koefisien silhouette untuk setiap titik data
26 silhouette_obj <- silhouette(cutree(hc, k = 2), dist(dataku))
27 # Plot koefisien silhouette
28 plot(silhouette_obj, main = "Silhouette Coefficient of Average Linkage")
29
30 # Membuat plot cluster silhouette
31 fviz_silhouette(silhouette_obj)
32
33 plot(hc, hang=-1, main = "Cluster Dendrogram")
34 n_clusters <- 2
35 # Dapatkan tinggi pemotongan untuk memisahkan kluster
36 cut_height <- hc$height[length(hc$height) - n_clusters + 2]
37
38 # Tambahkan garis untuk memisahkan kluster
39 rect.hclust(hc, k = n_clusters, border = c("red", "green"))
40
```



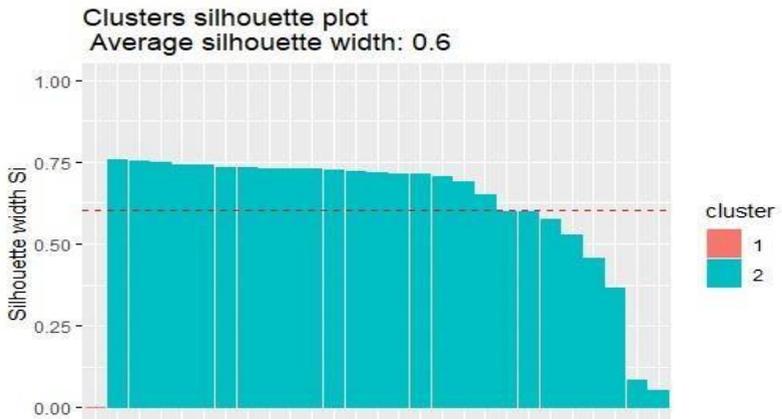
- Plot Silhouette Average Linkage



- Plot Gap Stat Average Linkage

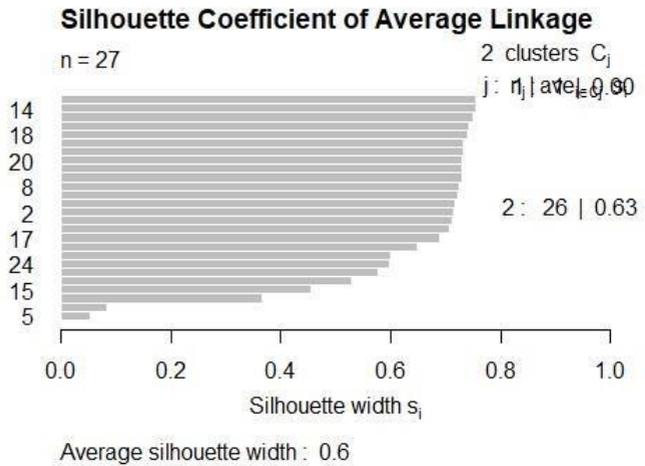


- Plot Cluster Silhouette Plot Average Silhouette Width

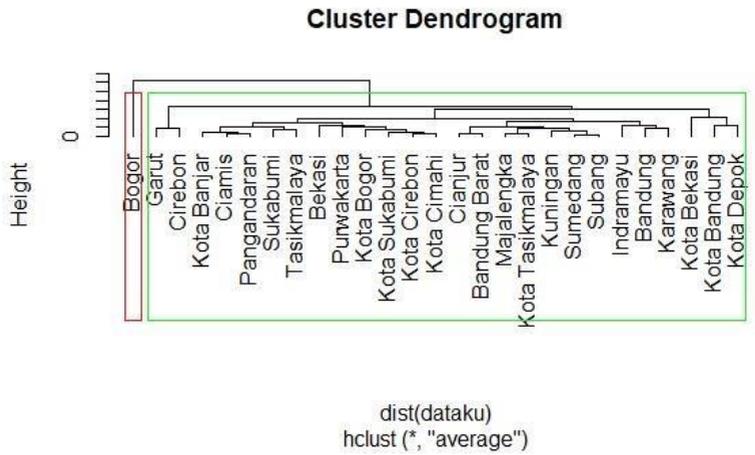




- Plot Silhouette Coefficient of Average Linkage



- Plot Dendrogram



RIWAYAT HIDUP

A. Identitas Diri

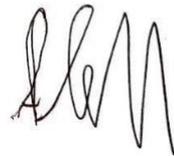
1. Nama : Nabila Garisoh Alfarisi
2. TTL : Demak, 15 Oktober 2000
3. Alamat Rumah : Perumnas Kalikondang Demak RT
05 RW 01 Blok E16
4. No. HP : 085714700553
5. Email : abilronaldo54@gmail.com

B. Riwayat Pendidikan

1. SDN Kalikondang 3 Demak
2. SMPN 2 Demak
3. SMAN 1 Demak

Semarang, 11 Oktober 2023

Pembuat Pernyataan



Nabila Garisoh Alfarisi

NIM. 1808046010