

Nama : Risang Ayu Siwi Eka Putri  
NIM : 221910888  
No Absen : 22  
Kelas : 3SD2  
Dosen : Farid Ridho, MT  
Mata Kuliah : Visualisasi Data dan Informasi  
Hari/Tanggal : Selasa, 21 Juni 2022

“SAYA MENYATAKAN BAHWA UJIAN INI SAYA KERJAKAN DENGAN JUJUR SESUAI KEMAMPUAN SAYA SENDIRI DAN TIDAK MENGUTIP SEBAGIAN ATAU SELURUH PEKERJAAN ORANG LAIN. JIKA SUATU SAAT DITEMUKAN SAYA MELANGGAR KETENTUAN UJIAN, SAYA SIAP MENERIMA KONSEKUENSI YANG BERLAKU.”



(Risang Ayu Siwi Eka Putri)

# Visualisasi Dan Analisis Clustering Provinsi di Indonesia Berdasarkan Indikator Kemiskinan Menggunakan Algoritma K-Means

Risang Ayu Siwi Eka Putri (221910888, 3SD2)

Dosen Pembimbing: Farid Ridho, MT

**Ringkasan—** Semakin meningkatnya jumlah penduduk, maka semakin menimbulkan berbagai permasalahan. Salah satunya adalah kemiskinan. Dalam pembangunan ekonomi, kemiskinan menjadi salah satu permasalahan yang dihadapi negara berkembang seperti Indonesia. Berdasarkan data dari BPS, angka kemiskinan di Indonesia pada Maret 2021 sebesar 10,14 persen. Angka tersebut meningkat sebanyak 0,36 persen dibandingkan dengan Maret 2020. Namun pada September 2021, angka kemiskinan menjadi 9,71 persen, terjadi penurunan sebanyak 0,43 persen dibandingkan Maret 2021. Upaya yang bisa dilakukan untuk pengentasan kemiskinan adalah pengelompokan wilayah berdasarkan karakteristik yang sama untuk mempermudah pemerintah dalam pengambilan keputusan. Sehingga dilakukan analisis cluster untuk mengelompokkan objek-objek sedemikian rupa sehingga objek dalam satu cluster sangat mirip dan objek diberbagai cluster cukup berbeda. Pengelompokan wilayah ini dilakukan berdasarkan karakteristik kemiskinan yaitu Garis Kemiskinan, Persentase Penduduk Miskin, Indeks Kedalaman Kemiskinan, dan Indeks Keparahan. Berdasarkan hasil analisis cluster diperoleh 3 cluster, yaitu cluster 1 (tingkat kemiskinan sedang) terdiri dari 16 provinsi, cluster 2 (tingkat kemiskinan rendah) terdiri dari 15 provinsi, dan cluster 3 (tingkat kemiskinan tinggi) terdiri dari 3 provinsi. Selain itu dilakukan visualisasi dari masing-masing indikator kemiskinan yang disatukan dalam bentuk dashboard.

**Kata Kunci—** Indikator, Kemiskinan, Cluster, K-Means

## I. LATAR BELAKANG

Semakin meningkatnya jumlah penduduk, maka semakin menimbulkan berbagai permasalahan. Salah satunya adalah kemiskinan. Kemiskinan merupakan ketidakmampuan seseorang atau kelompok untuk memenuhi hak-hak dasarnya dalam mempertahankan dan mengembangkan kehidupan yang bermartabat [8]. Dalam pembangunan ekonomi, kemiskinan menjadi salah satu permasalahan yang dihadapi negara berkembang seperti Indonesia. Masalah ini bersifat multidimensional dan kompleks sehingga menjadi prioritas dalam pembangunan.

Sejak awal tahun 2020, Indonesia terdampak pandemi covid-19 yang menyebabkan perekonomian menurun dan belum pulih hingga saat ini. Berdasarkan data dari Badan Pusat Statistik, angka kemiskinan di Indonesia pada Maret 2021 sebesar 10,14 persen. Angka tersebut meningkat sebanyak 0,36 persen poin dibandingkan dengan Maret 2020. Namun pada September 2021, angka kemiskinan menjadi 9,71 persen, terjadi penurunan sebanyak 0,43 persen poin dibandingkan Maret 2021.

Salah satu strategi pengentasan kemiskinan di Indonesia yang memiliki karakteristik wilayah yang berbeda yaitu diperlukan suatu kebijakan yang berbeda pula antar wilayah.

Upaya yang bisa dilakukan adalah pengelompokan wilayah berdasarkan karakteristik yang sama untuk mempermudah pemerintah dalam pengambilan keputusan. Salah satu metode pengelompokan wilayah ini adalah menggunakan analisis *cluster*.

Analisis *cluster* merupakan salah satu analisis peubah ganda yang digunakan untuk mengelompokkan objek-objek sedemikian rupa sehingga objek dalam satu cluster sangat mirip dan objek diberbagai cluster cukup berbeda [9]. Pengelompokan wilayah ini dilakukan berdasarkan karakteristik kemiskinan yaitu Garis Kemiskinan (GK), Persentase Penduduk Miskin, Indeks Kedalaman Kemiskinan (P1), dan Indeks Keparahan Kemiskinan (P2) [3].

Melalui pengelompokan ini, dapat dibentuk wilayah-wilayah yang bisa menjadi prioritas dalam penentuan kebijakan serta wilayah-wilayah yang dapat menjadi contoh dalam pengentasan kemiskinan. Sehingga tidak terjadi banyak ketimpangan antar wilayah dan angka kemiskinan Indonesia semakin menurun.

## II. TUJUAN PENELITIAN

Berdasarkan uraian pada latar belakang, maka penelitian ini bertujuan sebagai berikut

1. Mengetahui hasil analisis *cluster* provinsi di Indonesia berdasarkan indikator kemiskinan.
2. Memvisualisasikan masing-masing indikator kemiskinan dan hasil analisis *cluster* nya.
3. Mengetahui daerah yang menjadi prioritas pemerintah dalam membuat kebijakan agar tidak terjadi banyak ketimpangan antar wilayah dan menurunkan angka kemiskinan Indonesia.

## III. PENELITIAN TERKAIT

Achmad Bahaudin dkk (2021) melakukan penelitian yang berjudul “Analisis *Clustering* Provinsi Di Indonesia Berdasarkan tingkat Kemiskinan Menggunakan Algoritma *K-Means*”[4]. Penelitian tersebut menggunakan metode clustering dengan algoritma K-Means untuk mengelompokkan provinsi berdasarkan tingkat kemiskinannya. Pengolahan datanya menggunakan software WEKA. Hasil clustering menunjukkan bahwa terbentuk 3 kelompok dengan cluster 0 (Memiliki tingkat kemiskinan rendah) terdiri dari 9 provinsi, cluster 1 (Memiliki tingkat kemiskinan sedang) terdiri dari 21 provinsi, dan cluster 2 (Memiliki tingkat kemiskinan tinggi) terdiri dari 3 provinsi.

Nugroho dan Nico (2018) melakukan penelitian yang berjudul “Analisis *Clustering K-Means* Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018”[6]. Penelitian tersebut menggunakan metode *K-Means Clustering* untuk

mengelompokkan kemiskinan di Jawa Barat berdasarkan 13 indikator kemiskinan yang berada pada sektor ekonomi, sumber daya manusia, pendidikan, dan kesehatan. Hasil *clustering* menunjukkan bahwa terbentuk 5 kelompok, dimana *cluster* ke-1 menjadi prioritas dalam mempertimbangkan kebijakan untuk meningkatkan kesejahteraan penduduk.

Debrina (2018) melakukan penelitian yang berjudul “Analisis Tingkat Kemiskinan Indonesia” [7]. Penelitian ini menggunakan kombinasi kedua metode yaitu metode hirarki kemudian dilanjutkan dengan metode non hirarki untuk mengelompokkan tingkat kemiskinan berdasarkan 3 indikator. Hasil *clustering* menunjukkan bahwa terbentuk 3 kelompok dengan *cluster* 1 provinsi miskin terdiri dari 3 provinsi, *cluster* 2 provinsi sedang terdiri dari 13 provinsi, dan *cluster* 3 provinsi tidak miskin terdiri dari 18.

Selain menggunakan *K-Means*, terdapat juga penelitian terkait yang menggunakan *K-Medoid*. Disty (2019) mengelompokkan garis kemiskinan yang menghasilkan 2 kelompok yaitu tingkat kemiskinan tinggi 23 provinsi dan tingkat kemiskinan rendah 11 provinsi [11]. Adina (2021) mengelompokkan kemiskinan dan indeks pembangunan yang menghasilkan 3 kelompok, yaitu kemiskinan tinggi 10 provinsi, kemiskinan sedang 19 provinsi, dan kemiskinan rendah 5 provinsi [2]. Gifthera (2021) mengelompokkan persentase penduduk miskin tahun 2007-2019 yang menghasilkan 3 kelompok, yaitu kemiskinan tinggi terdiri dari 3 provinsi, kemiskinan sedang terdiri dari 12 provinsi, dan kemiskinan rendah terdiri dari 19 provinsi [8].

#### IV. KAJIAN PUSTAKA

##### A. Kemiskinan

Kemiskinan secara etimologis berasal dari kata “miskin”, yang berarti tidak berharta benda dan serba kekurangan. BPS mendefinisikan kemiskinan sebagai ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar baik makanan ataupun bukan makanan yang diukur dari sisi pengeluaran. Dikategorikan sebagai penduduk miskin apabila memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan.

##### B. Garis Kemiskinan

BPS menyebutkan bahwa Garis Kemiskinan (GK) merupakan nilai rupiah pengeluaran minimum yang diperlukan oleh seseorang dalam memenuhi kebutuhan pokok hidupnya selama satu bulan, baik kebutuhan makanan maupun non makanan. Garis Kemiskinan Makanan adalah nilai pengeluaran minimum untuk kebutuhan makanan yang disetarakan dengan 2100 kilokalori per kapita per hari nya yang diwakili oleh 52 jenis komoditi kebutuhan dasar makanan.

Sedangkan Garis Kemiskinan Non Makanan adalah nilai pengeluaran minimum untuk kebutuhan dasar yang berupa perumahan, sandang, pendidikan, dan kesehatan dengan diwakili oleh 51 jenis komoditi di perkotaan dan 47 jenis komoditi di pedesaan. Garis Kemiskinan dirumuskan sebagai berikut.

$$GK = GKM + GKNM \quad (1)$$

GK : Garis Kemiskinan

GKM : Garis Kemiskinan Makanan

GKNM : Garis Kemiskinan Non Makanan

##### C. Persentase Penduduk Miskin

Persentase penduduk miskin merupakan persentase penduduk yang berada di bawah garis kemiskinan [7]. Secara sederhana, Head Count Index (HCI-PO) mengukur proporsi yang dikategorikan miskin. Persentase penduduk miskin yang tinggi mengartikan bahwa tingkat kemiskinan suatu wilayah tersebut tinggi dan sebaliknya. HCI-PO dapat dirumuskan sebagai berikut.

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[ \frac{z - y_i}{z} \right]^{\alpha} \quad (2)$$

$P_{\alpha}$  : persentase penduduk miskin

$\alpha$  : 0

$z$  : garis kemiskinan

$y_i$  : rata-rata pengeluaran per kapita sebulan penduduk yang berada di bawah garis kemiskinan

$q$  : banyaknya penduduk yang berada di bawah garis kemiskinan

$n$  : jumlah penduduk

##### D. Indeks Kedalaman Kemiskinan

Indeks Kedalaman Kemiskinan (*Poverty Gap Indeks - P1*) adalah ukuran rata-rata kesenjangan pengeluaran pada masing-masing penduduk miskin terhadap garis kemiskinan. Nilai agregat dari indeks ini menunjukkan bahwa biaya mengentaskan kemiskinan dengan membuat target transfer yang sempurna terhadap penduduk miskin dalam hal tidak terdapat biaya transaksi dan factor penghambat [7]. Penurunan pada nilai indeks kedalaman menunjukkan bahwa rata-rata pengeluaran penduduk miskin cenderung mendekati garis kemiskinan dan ketimpangan pengeluaran penduduk miskin semakin menyempit. Indeks Kedalaman dapat dirumuskan sebagai berikut.

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[ \frac{z - y_i}{z} \right]^{\alpha} \quad (3)$$

$P_{\alpha}$  : indeks kedalaman kemiskinan

$\alpha$  : 1

$z$  : garis kemiskinan

$y_i$  : rata-rata pengeluaran per kapita sebulan penduduk yang berada di bawah garis kemiskinan

$q$  : banyaknya penduduk yang berada di bawah garis kemiskinan

$n$  : jumlah penduduk

##### E. Indeks Keparahan Kemiskinan

Indeks Keparahan Kemiskinan (*Poverty Severity Indeks - P2*) memberikan suatu informasi mengenai gambaran penyebaran pengeluaran diantara penduduk miskin [7]. Semakin tinggi nilai indeks nya, maka semakin tinggi pula

ketimpangan pengeluaran diantara penduduk miskin. Indeks Keparahan dapat dirumuskan sebagai berikut.

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[ \frac{z-y_i}{z} \right]^{\alpha} \quad (4)$$

$P_{\alpha}$  : indeks keparahan kemiskinan  
 $\alpha$  : 2  
 $z$  : garis kemiskinan  
 $y_i$  : rata-rata pengeluaran per kapita sebulan penduduk yang berada di bawah garis kemiskinan  
 $q$  : banyaknya penduduk yang berada di bawah garis kemiskinan  
 $n$  : jumlah penduduk

#### F. Analisis Cluster

Analisis *cluster* merupakan suatu metode dalam statistik yang bertujuan untuk mengelompokkan objek ke dalam suatu kelompok yang memiliki kemiripan yang tinggi dibandingkan dengan objek yang berada dalam kelompok lain [10]. *Clustering* menempatkan suatu objek yang memiliki kesamaan atau jarak yang berdekatan ke dalam suatu kelompok dan membuat jarak antar kelompok sejauh mungkin. Metode ini adalah salah satu metode pada *data mining* yang bersifat tanpa arahan (*unsupervised*) yang artinya diterapkan tanpa adanya latihan (*training*) dan tanpa guru serta tidak memerlukan target output.

#### G. Algoritma K-Means

Algoritma *K-Means* merupakan salah satu metode analisis *cluster* yang non hierarki. Metode ini paling sering digunakan dan lebih mudah diimplementasikan. Pada *K-Means*, biasanya peneliti terlebih dahulu menentukan banyak *cluster* yang akan dibentuk [10]. Dalam metode ini, digunakan ukuran kemiripan untuk mengelompokkan suatu objek atau dalam konsep jarak [4]. Apabila jarak dua objek cukup dekat, maka semakin tinggi kemiripan yang akan menjadikan dalam satu kelompok. Sebaliknya, semakin jauh jaraknya, maka semakin tinggi ketidakmiripannya.

Untuk menentukan jarak pada algoritma ini menggunakan jarak *Euclidean* yang dirumuskan sebagai berikut.

$$D(ij) = \sqrt{\sum_{k=1}^p |X_{ki} - X_{kj}|^2} \quad (5)$$

$D(ij)$  : jarak data ke (i) ke pusat *cluster* (j)  
 $X_{ki}$  : data ke (i) pada atribut data ke (k)  
 $X_{kj}$  : titik pusat (j) pada atribut (k)

### V. METODE PENELITIAN

#### A. Sumber Data

Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik. Data yang digunakan merupakan data-data indikator kemiskinan yang diambil pada bulan September 2021. Variabel yang digunakan sebagai indikator adalah Garis Kemiskinan, Persentase Penduduk Miskin, Indeks Kedalaman Kemiskinan, dan Indeks Keparahan Kemiskinan.

#### B. Preprocessing Data

Langkah pertama yang dilakukan sebelum masuk ke clustering adalah preprocessing pada data. Preprocessing data merupakan teknik awal data mining untuk mengubah data mentah atau yang biasa disebut raw data menjadi informasi yang lebih bersih dan bisa digunakan untuk pengolahan selanjutnya. Beberapa masalah umum yang diselesaikan dalam tahap ini adalah mengenai missing value, data noise, dan data yang tidak konsisten. Selain itu dilakukan juga standarisasi data dengan transformasi. Standarisasi digunakan apabila terdapat perbedaan ukuran satuan yang besar antara variabel-variabel yang diteliti [10]. Tahap ini sangat penting karena kesalahan dan redundansi pada data yang belum terselesaikan akan menyebabkan berkurangnya akurasi hasil analisis

#### C. Analisis Data

Setelah melakukan *preprocessing* pada data, dilakukan analisis data dengan berbagai tahap agar memiliki hasil dan bisa ditarik kesimpulan. Adapun tahapan analisis yang dilakukan adalah sebagai berikut.

##### 1) Analisis Deskriptif

Menurut Sugiyono (2017:147) analisis deskriptif adalah statistik yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul sebagaimana adanya tanpa bermaksud membuat kesimpulan yang berlaku umum atau generalisasi. Analisis deskriptif pada penelitian ini antara lain mean, median, modus, nilai minimum, nilai maksimum. Analisis ini bisa digunakan untuk melihat provinsi yang memiliki indikator kemiskinan terendah maupun tertinggi.

##### 2) Analisis cluster menggunakan K-Means

Tahapan dalam algoritma *K-Means* adalah sebagai berikut [6].

- Menentukan nilai k sebagai jumlah *cluster* yang akan dibentuk
- Menentukan *centroid* atau titik pusat awal secara random
- Menghitung setiap observasi ke *centroid* menggunakan jarak Euclidean jika tidak terjadi korelasi. Jika terjadi korelasi maka menggunakan jarak mahalanobis
- Mengelompokkan setiap observasi berdasarkan jarak terdekat dengan *centroid*
- Melakukan iterasi dari poin ke-ii sampai akhir dengan menentukan nilai *centroid*. Hingga nilai tersebut tetap dan anggota *cluster* tidak berpindah ke *cluster* lain.

#### D. Visualisasi Data

Visualisasi data merupakan representasi visual dari data yang dapat mempermudah penyampaian informasi secara efektif. Manusia memiliki kemampuan yang lebih baik untuk menganalisis sejumlah data yang direpresentasikan secara visual daripada berbentuk tulisan. Sehingga penting untuk

memilih visual yang tepat agar informasi dapat tersampaikan dengan baik bagi pembaca.

Visualisasi data yang dilakukan merupakan visualisasi interaktif terhadap data indikator kemiskinan. Masing-masing indikator akan dibuat bentuk visualisasinya yang kemudian disatukan dalam dashboard. *Software* yang digunakan untuk melakukan visualisasi dalam penelitian ini adalah Tableau. *Software* ini merupakan salah satu *tools* yang digunakan untuk mengolah data menjadi sebuah visual interaktif yang menarik.

## VI. HASIL DAN PEMBAHASAN

### A. Pengumpulan Data

Data yang dikumpulkan dari BPS diperoleh 34 provinsi di Indonesia dengan 4 variabel. Variabel tersebut menjadi indikator kemiskinan yang akan dilakukan analisis *cluster*. Indikator yang digunakan adalah Garis Kemiskinan, Persentase Penduduk Miskin, Indeks Kedalaman Kemiskinan, dan Indeks Keparahan Kemiskinan.

### B. Preprocessing Data

Pada preprocessing data dilakukan data *cleaning*, data *integration*, dan data *transformation*. Data *cleaning* berfungsi untuk mengganti missing value, mengatasi data noise, dan menghilangkan data yang redundan. Data *integration* berfungsi menggabungkan variabel-variabel menjadi satu bentuk data. Data *transformation* berfungsi untuk menormalisasikan data, pada tahap ini data distandarisasi agar tidak terdapat perbedaan satuan yang besar. Berikut contoh data yang telah dilakukan *preprocessing* sampai menstandarisasi data.

TABEL I  
CONTOH DATA INDIKATOR KEMISKINAN SETELAH PREPROCESSING

No	Provinsi	Garis Kemiskinan	Persentase Penduduk Miskin	Indeks Kedalaman	Indeks Keparahan
1	Aceh	0.387	0.943	0.731	0.579
2	Sumatera Utara	0.194	-0.358	-0.334	-0.287
3	Sumatera Barat	0.674	-0.811	-0.681	-0.589
4	Riau	0.513	-0.633	-0.589	-0.488
5	Jambi	0.005	-0.510	-0.589	-0.589

### C. Analisis Data

Pada analisis data dilakukan tahapan sebagai berikut.

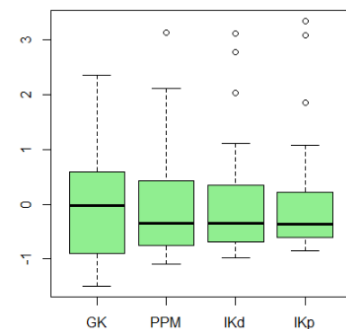
#### 1) Analisis Deskriptif

Pada analisis deskriptif dihitung nilai minimum, nilai maksimum, rata-rata dan standar deviasi untuk melihat persebaran dari data. Berikut ini merupakan tabel analisis deskriptif dari data indikator kemiskinan menggunakan data sebelum standarisasi.

TABEL II  
ANALISIS DESKRIPTIF DATA INDIKATOR KEMISKINAN

No	Indikator	Minimum	Rata-Rata	Maksimum	Standar Deviasi
1	Garis Kemiskinan	287467	391339	555279	69720.36
2	Persentase Penduduk Miskin	4.56	10.43	27.38	5.41
3	Indeks Kedalaman	0.55	1.92	6.31	1.41
4	Indeks Keparahan	0.10	0.52	2.18	0.50

Selain itu, dilakukan visualisasi berupa boxplot menggunakan data yang telah distandarisasi untuk mengetahui outlier pada data.



Gambar 1. Boxplot Data Indikator Kemiskinan

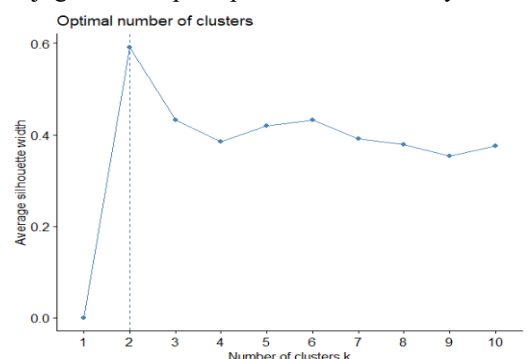
Dari boxplot di atas, dapat diketahui bahwa indikator Persentase Penduduk Miskin terdapat 1 outlier, indikator Indeks Kedalaman terdapat 3 outlier, dan indikator Indeks Keparahan terdapat 3 outlier.

#### 2) Analisis *cluster* menggunakan *K-Means*

Analisis *cluster* dilakukan dengan beberapa tahapan. Berikut ini adalah tahapan-tahapan menggunakan algoritma *K-Means*

##### i. Menentukan nilai K

Untuk menentukan nilai K, menggunakan metode *Silhouette* sebagai pertimbangan. Selain itu juga melihat pada penelitian sebelumnya.



Gambar 2. K optimal dengan metode *Silhouette*



Dari gambar di atas dapat dilihat bahwa K optimal menggunakan metode *Silhouette* menghasilkan 2 *cluster*. Namun, pada penelitian ini juga mempertimbangkan penelitian sebelumnya dan mempertimbangkan kemudahan dalam interpretasi. Sehingga nilai K yang digunakan adalah 3 *cluster*. *Cluster* yang terbentuk adalah tingkat kemiskinan tinggi, sedang, dan rendah.

#### ii. *Cluster* menggunakan *K-Means*

Berdasarkan hasil dari analisis *cluster* menggunakan *K-Means* terbentuk 3 *cluster*. *Cluster* 1 terdiri dari 16 provinsi dengan rata-rata Garis Kemiskinan sebesar 348168 rupiah, Persentase Penduduk miskin sebesar 12.1%, Indeks Kedalaman sebesar 2.15 dan Indeks Keparahan sebesar 0.552. provinsi yang termasuk dalam *cluster* 1 adalah Aceh, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Jawa Tengah, D.I Yogyakarta, Jawa Timur, Nusa Tenggara Barat, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, dan Maluku.

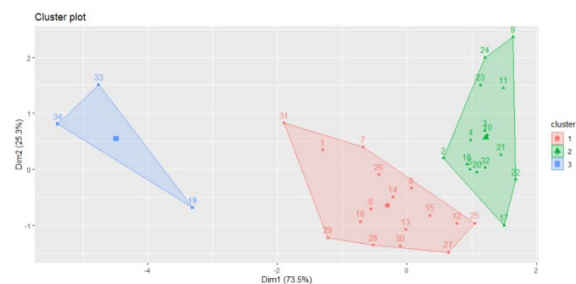
*Cluster* 2 terdiri dari 15 provinsi dengan rata-rata Garis Kemiskinan sebesar 429798 rupiah, Persentase Penduduk miskin sebesar 6.1%, Indeks Kedalaman sebesar 0.93 dan Indeks Keparahan sebesar 0.217. provinsi yang termasuk dalam *cluster* 2 adalah Sumatera Utara, Sumatera Barat, Riau, Jambi, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Banten, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, dan Maluku Utara.

*Cluster* 3 terdiri dari 3 provinsi dengan rata-rata Garis Kemiskinan sebesar 429293 rupiah, Persentase Penduduk miskin sebesar 23.2%, Indeks Kedalaman sebesar 5.65 dan Indeks Keparahan sebesar 1.89. provinsi yang termasuk dalam *cluster* 3 adalah Nusa Tenggara Timur, Papua, dan Papua Barat.

Dari hasil *cluster* yang terbentuk, dihitung pula ukuran evaluasi *cluster* atau uji validasi untuk melihat kesesuaian model yang digunakan untuk pengelompokan. Uji validasi ini menggunakan *R-Square*. Didapat bahwa *R-Square* sebesar 0.732 atau 73.2%. Artinya model yang dibuat untuk pengelompokkan sudah cukup baik.

#### iii. Plot hasil analisis *cluster*

Berikut ini merupakan plot hasil *cluster* yang membentuk 3 kelompok tingkat kemiskinan pada 34 provinsi di Indonesia.



Gambar 3. Plot hasil analisis *cluster*

#### iv. Karakteristik Setiap *cluster*

Berikut ini merupakan tabel rata-rata setiap *cluster*. Dari rata-rata tersebut akan diambil kesimpulan mengenai tingkat kemiskinan mana yang tinggi, sedang, dan rendah.

TABEL III  
RATA-RATA INDIKATOR SETIAP *CLUSTER*

Cluster	Garis Kemiskinan	Persentase Penduduk Miskin	Indeks Kedalaman	Indeks Keparahan
1	348168	12.1	2.15	0.55
2	429798	6.1	0.93	0.22
3	429293	23.2	5.65	1.89

*Cluster* 1 adalah kelompok provinsi yang memiliki tingkat kemiskinan sedang, dimana rata-rata indikator Persentase Penduduk Miskin, Indeks Kedalaman, dan Indeks Keparahan sedang. Namun Indikator Garis Kemiskinan memiliki rata-rata yang rendah.

*Cluster* 2 adalah kelompok provinsi yang memiliki tingkat kemiskinan rendah, dimana rata-rata indikator Persentase Penduduk Miskin, Indeks Kedalaman, dan Indeks Keparahan rendah. Namun Indikator Garis Kemiskinan memiliki rata-rata yang tinggi.

*Cluster* 3 adalah kelompok provinsi yang memiliki tingkat kemiskinan tinggi, dimana rata-rata indikator Persentase Penduduk Miskin, Indeks Kedalaman, dan Indeks Keparahan tinggi. Namun Indikator Garis Kemiskinan memiliki rata-rata yang sedang.

Dari hasil *cluster*, dapat diketahui bahwa daerah yang sebaiknya menjadi prioritas pemerintah dalam membuat kebijakan agar tidak terjadi banyak ketimpangan antar wilayah dan menurunkan angka kemiskinan adalah wilayah *cluster* 3 yang memiliki tingkat kemiskinan tinggi, yaitu provinsi Nusa Tenggara Timur, Papua, dan Papua Barat.

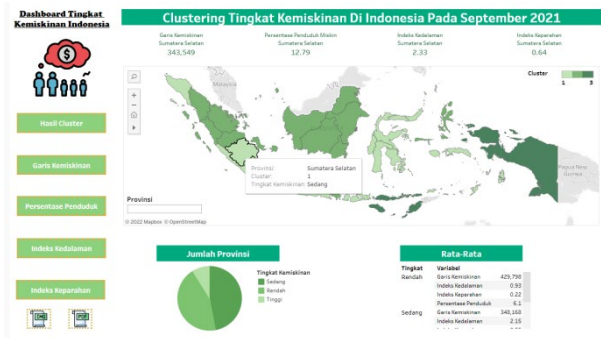
#### D. Visualisasi Data

Pada tahap ini, dilakukan visualisasi data pada setiap indikator tingkat kemiskinan dan visualisasi hasil dari analisis *cluster*. Selain itu visualisasi data akan disatukan ke dalam

bentuk dashboard. Visualisasi yang digunakan adalah peta tematik, *bar chart*, *pie chart*, *boxplot*, dan dalam bentuk tabulasi data.

Dashboard yang dibuat memiliki beberapa fitur yang memudahkan pengguna mengaksesnya. Tombol navigasi yang terletak dalam sidebar sebelah kiri yang mengarahkan untuk ke halaman yang dipilih. *Search Provinsi* yang digunakan untuk mencari provinsi yang dituju. *Action* yaitu suatu aksi dimana ketika mengarahkan kursor atau mengklik suatu provinsi akan terlihat keterangan masing-masing variabelnya. *Download* yaitu untuk mengunduh visualisasi dalam bentuk *image* ataupun pdf. *Tooltip* yaitu ketika mengarahkan kursor ke visualisasi nya maka akan tertera penjelasan singkat mengenai hal yang ditunjuk. Pada penelitian ini, terdapat 5 halaman dashboard, yaitu

### 1) Hasil Cluster



Gambar 5. Halaman dashboard hasil cluster (1)



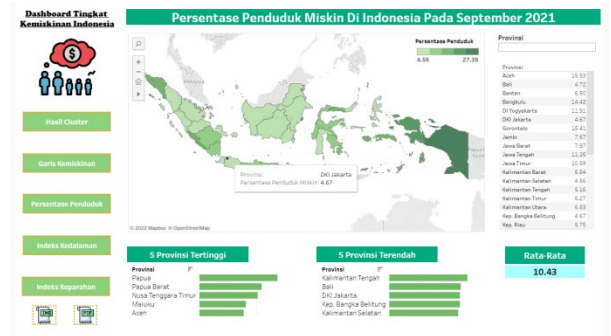
Gambar 6. Halaman dashboard hasil cluster (2)

### 2) Garis Kemiskinan



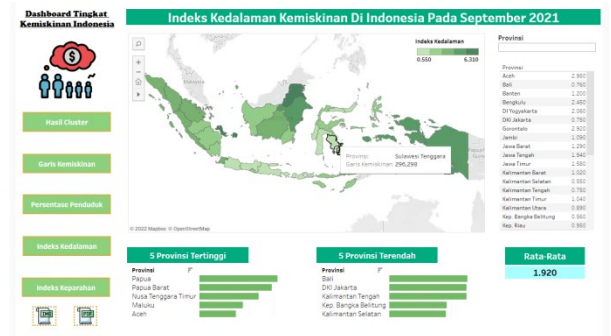
Gambar 7. Halaman dashboard garis kemiskinan

### 3) Persentase Penduduk Miskin



Gambar 8. Halaman dashboard persentase penduduk miskin

### 4) Indeks Kedalaman Kemiskinan



Gambar 9. Halaman dashboard indeks kedalaman kemiskinan

### 5) Indeks Keparahannya Kemiskinan



Gambar 10. Halaman dashboard indeks keparahannya kemiskinan

Link Github :

<https://github.com/risangayusiwi/Dashboard-Visdat>

Link Dashboard :

[https://public.tableau.com/app/profile/risang.ayu.siwi/viz/Dashboard\\_16553849736090/DBCL?publish=yes](https://public.tableau.com/app/profile/risang.ayu.siwi/viz/Dashboard_16553849736090/DBCL?publish=yes)

## VII. KESIMPULAN

Berdasarkan hasil penelitian analisis *cluster* menggunakan *K-Means* didapatkan 3 *cluster* provinsi di Indonesia berdasarkan tingkat kemiskinan. *Cluster 1* adalah kelompok provinsi yang memiliki tingkat kemiskinan sedang yang terdiri dari 16 provinsi, yaitu Aceh, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Jawa Tengah, D.I Yogyakarta, Jawa Timur, Nusa Tenggara Barat, Sulawesi

Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, dan Maluku.

*Cluster 2* adalah kelompok provinsi yang memiliki tingkat kemiskinan rendah yang terdiri dari 15 provinsi, yaitu Sumatera Utara, Sumatera Barat, Riau, Jambi, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Banten, Bali, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, dan Maluku Utara. *Cluster 3* adalah kelompok provinsi yang memiliki tingkat kemiskinan tinggi yang terdiri dari 3 provinsi, yaitu Nusa Tenggara Timur, Papua, dan Papua Barat. Selain itu wilayah di *cluster 3* menjadi prioritas pemerintah dalam membuat kebijakan agar tidak terjadi banyak ketimpangan antar wilayah dan menurunkan angka kemiskinan di Indonesia.

Hasil uji validasi menggunakan *R-Square* didapat sebesar 73.2%, artinya model yang dibuat untuk mengelompokkan sudah cukup baik.

#### DAFTAR PUSTAKA

- [1] Abidin, Z. (2018). Pengelompokan Kabupaten/Kota Di Jawa Timur Berdasarkan Indikator Kemiskinan Dengan Menggunakan Analisis Cluster Hierarki (Doctoral dissertation, Institut Teknologi Sepuluh Nopember).
- [2] Astasia, A. (2021). Analisis Cluster Kemiskinan Dan Indeks Pembangunan Manusia di Indonesia dengan Data Outlier Menggunakan K-Medoids dan Analisis Biplot. *Media Edukasi Data Ilmiah dan Analisis (MEDIAN)*, 4(02), 1-8.
- [3] Badan Pusat Statistik (2021). Penghitungan dan Analisis Kemiskinan Makro Indonesia Tahun 2021
- [4] Bahauddin, A., Fatmawati, A., & Sari, F. P. (2021). Analisis Clustering Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means. *Jurnal Manajemen Informatika Dan Sistem Informasi*, 4(1), 1-8.
- [5] Dqlab.id. (2021). Pentingnya Preprocessing Dalam Pengolahan Data Statistik. Diakses pada 31 Mei 2022. <https://www.dqlab.id/pentingnya-preprocessingdalam-pengolahan-data-statistik>
- [6] Febianto, N. I., & Palasara, N. (2019). Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 8(2), 130-140.
- [7] Ferezagia, D. V. (2018). Analisis tingkat kemiskinan di Indonesia. *Jurnal Sosial Humaniora Terapan*, 1(1).
- [8] Lestari, G. D. (2021). Analisis Clustering menggunakan K-Medoid pada Data Penduduk Miskin Indonesia. *JURSIMA (Jurnal Sistem Informasi dan Manajemen)*, 9(3), 282-290.
- [9] Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. Yogyakarta: Graha Ilmu, 978(979), 756.
- [10] Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia. *J. Mat."MANTIK*, 4(1), 22-31.
- [11] Wahyuli, D., Handrizal, H., Parlina, I., Windarto, A. P., Suhendro, D., & Wanto, A. (2019). Mengelompokkan Garis Kemiskinan Menurut Provinsi Menggunakan Algoritma K-Medoids. In *Prosiding Seminar Nasional Riset Information Science (SENARIS) (Vol. 1, pp. 452-461)*.