

SEGMENTASI PENDUDUK MISKIN DI INDONESIA MENGGUNAKAN ALGORITMA *K-MEANS*

Agita Vidiasti Rivallinata¹⁾, Tubagus Mohammad Akhriza²⁾, Dwi Safiroh Utsalina³⁾
Sistem Informasi, STMIK PPKIA Pradnya Paramita
agitavidiastrivallinata@gmail.com, akhriza@stimata.ac.id, utsalina@stimata.ac.id

Abstract

The Covid-19 pandemic has led to an increase in the number of poor people in Indonesia due to government policies aimed at curbing the virus's spread. Grouping poverty levels in Indonesia is crucial for policy-making. This study aims to classify poverty data in Indonesia based on attributes such as GKM, GKNM, IkdK, and IKpK, utilizing the K-Means algorithm for data mining. The results reveal three poverty clusters: low, medium, and high. Clustering the data before, during, and after the pandemic without employing binning techniques did not lead to cluster shifts. However, using binning resulted in cluster shifts in certain provinces in 2020 and 2022. Clustering the data during the pandemic peak, towards normalcy, without binning, showed a shift from high to low poverty levels in Maluku and East Nusa Tenggara provinces. On the other hand, applying binning led to a shift from high to low poverty in East Nusa Tenggara province and from medium to low poverty in Bengkulu province. The Silhouette Coefficient, used as an evaluation metric, ranged from 0.54 to 0.59, indicating that the formed clusters have a good interpretation and are close to 1.

Keywords: *K-Means, Clustering, Silhouette Coefficient*

1. PENDAHULUAN

Masa pandemi Covid-19 yang menyerang Indonesia di awal tahun 2020 mengakibatkan pemerintah melakukan tindakan dengan mengeluarkan berbagai kebijakan untuk menekan penyebarannya. Kebijakan tersebut berdampak pada perekonomian masyarakat Indonesia menjadi terhambat serta memicu peningkatan tingkat pengangguran dan kemiskinan. Jumlah penduduk miskin mengalami peningkatan dibandingkan sebelum adanya pandemi sekitar 1,16 juta hingga 3,78 juta orang [1].

Berkaitan dengan kondisi kemiskinan tersebut, sudah seharusnya pemerintah melakukan upaya untuk mengurangi jumlah penduduk miskin di setiap provinsi melalui berbagai program bantuan. Diperlukan sebuah langkah pemetaan bagi tiap provinsi sesuai dengan tingkat garis kemiskinan sehingga dapat diambil sebuah solusi yang tepat sesuai dengan kelompoknya [2]. Penanggulangan kemiskinan dapat didukung dengan salah satu strategi seperti adanya data yang akurat

[3]. Oleh karena itu, penting bagi pemerintah melihat data aktual dan dibandingkan dengan data pada tahun sebelumnya untuk mengambil solusi terbaik mengatasi kemiskinan.

Menurut Badan Pusat Statistik (BPS), kemiskinan adalah ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran per kapita per bulan. Penduduk dikategorikan sebagai penduduk miskin apabila memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan. Garis kemiskinan makanan merupakan nilai pengeluaran kebutuhan minimum pada makanan dan disetarakan dengan 2100 kilo kalori per kapita per hari. Garis kemiskinan non makanan merupakan nilai pengeluaran minimum untuk kebutuhan non-makanan berupa perumahan, sandang, pendidikan dan kesehatan [4].

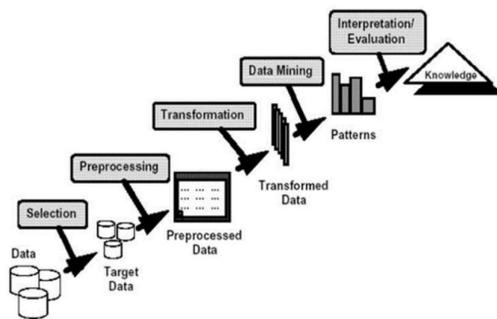
Berdasarkan latar belakang diatas, penelitian ini bertujuan melakukan segmentasi penduduk miskin di Indonesia berdasarkan atribut Garis Kemiskinan

Makanan (GKM), Garis Kemiskinan Non Makanan (GKNM), Indeks Kedalaman Kemiskinan (IkdK) dan Indeks Keparahan Kemiskinan (IkpK) menggunakan algoritma *K-Means*. Untuk mengetahui kualitas *clustering* yang dihasilkan menggunakan metode *Silhouette Coefficient* yang juga menjadi acuan beberapa penelitian. Serta mengetahui perubahan *cluster* kemiskinan dari tahun 2021 semester ke-2 (September) dan 2022 semester ke-1 (Maret) dengan tujuan untuk menjelaskan terjadi atau tidaknya perubahan segmen kemiskinan dari puncak masa pandemi pada akhir 2021 menuju kehidupan kenormalan baru di awal 2022.

2. KAJIAN LITERATUR

2.1 Data Mining

Data *mining* adalah sebuah teknik untuk mengidentifikasi suatu informasi secara menyeluruh terkait pada berbagai *dataset* yang besar untuk mendapat pengetahuan baru [3]. Data *mining* bermanfaat untuk mengolah data mentah menjadi kumpulan informasi yang dapat dijadikan sebagai penunjang keputusan secara efektif [5]. Menurut [6] dalam data *mining* terdapat beberapa proses seperti pada Gambar 1.



Gambar 1 KDD

Berikut adalah penjelasan tahapan data *mining* berdasarkan Gambar 1.

1. Data *Selection* yaitu menciptakan himpunan data target dengan memilih himpunan data yang kemudian hasil seleksi disimpan dalam suatu berkas yang terpisah dari basis data operasional.
2. *Pre-processing* yaitu penghapusan noise dengan membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data.

3. *Transformation* yaitu proses integrasi pada data yang telah dipilih sehingga data sesuai untuk proses data *mining*.
4. Data *Mining* yaitu pemilihan teknik, metode atau algoritma yang tepat dan sangat bergantung pada tujuan dan proses KDD secara keseluruhan.
5. *Interpretation* atau *Evaluation* merupakan tahap pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

2.2 K-Means Clustering

K-Means Clustering merupakan salah satu metode pengelompokan data nonhierarki (selatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain [7]. *Flowchart* dari algoritma *K-Means* seperti pada Gambar 2.



Gambar 2 *Flowchart* Algoritma *K-Means*

Langkah-langkah dalam menggunakan algoritma *K-Means* menurut Merliana, Ernawati & Santoso sebagai berikut:

1. Tentukan k sebagai jumlah *cluster* yang akan dibentuk.
2. Tentukan k *Centroid* (titik pusat *cluster*) awal secara random/acak.

$$v = \frac{\sum_{i=0}^n x_i}{n} ; i = 1,2,3, \dots n$$

Dimana;

v = centroid pada cluster

x_i = objek ke-1

n = jumlah objek anggota cluster

3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing cluster. Untuk menghitung jarak antara objek dengan centroid dapat menggunakan Euclidian Distance.

$$d(x, y) = |x - y| = \sqrt{\sum_{i=0}^n (x_i - y_i)^2} ; i = 1, 2, 3, \dots, n$$

Dimana;

x_i = objek x ke-i

y_i = daya y ke-i

n = banyaknya objek

4. Alokasikan masing-masing objek ke dalam centroid yang paling dekat.
5. Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan (1).
6. Ulangi langkah 3 jika posisi centroid baru tidak sama.

2.3 Metode Elbow

Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah cluster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah cluster yang akan membentuk siku pada suatu titik [8]. Metode Elbow digunakan untuk menentukan jumlah cluster dari dataset sehingga total dalam variasi cluster dapat diminimalkan, penentuan jumlah cluster dimulai dari $k=2$ dan terus bertambah di setiap langkah dengan ditambah 1 pada nilai k [9]. Berikut rumus untuk mendapatkan perbandingannya dengan menghitung SSE (Sum Of Square Error) dari setiap nilai cluster.

$$SSE = \sum_{K=1}^K \sum_{X_i \in S_K} \|X_i C_k\|_2^2$$

Keterangan:

k = jumlah cluster

X_i = nilai atribut data ke-i

C_k = nilai atribut titik pusat cluster ke-i

Langkah-langkah metode Elbow untuk menentukan nilai k sebagai berikut:

1. Inisialisasi awal nilai k .
2. Menaikkan nilai k .

3. Menghitung hasil sum of square error dari setiap nilai k .
4. Melihat hasil sum of square error dari nilai k yang turun secara drastis.
5. Menetapkan nilai k yang berbentuk siku.

2.4 Metode Silhouette Coefficient

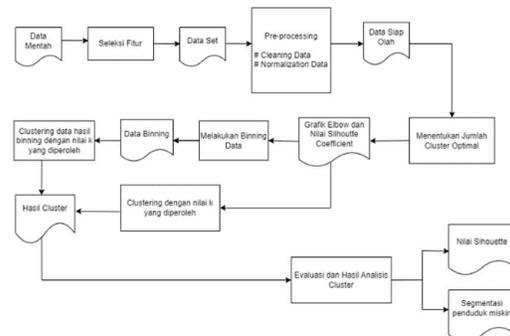
Silhouette Coefficient digunakan pada tahap evaluasi ini dengan fungsi untuk menguji kualitas dari cluster yang telah dihasilkan. Selanjutnya dapat digunakan mengambil keputusan untuk menentukan digunakan atau tidaknya hasil dari data mining tersebut. Silhouette Coefficient memiliki nilai dengan rentang -1 hingga 1, dengan ketentuan cluster yang semakin baik ketika nilai dari rata-rata yang dihasilkan dekat dengan rentang nilai 1 dan buruk apabila dekat dengan rentang -1. Rentang nilai ini dipergunakan untuk menunjukkan korelasi kemiripan dari data yang telah dikelompokkan ke dalam cluster [3]. Interpretasi subjektif dari koefisien silhouette yang didefinisikan sebagai lebar silhouette rata-rata maksimal untuk seluruh kumpulan data ditunjukkan pada Tabel 1 [10].

Tabel 1 Pengukuran Koefisien Silhouette
Sumber : [10]

Koefisien Silhouette	Interpretasi yang diusulkan
0,71-1	Struktur kuat
0,51-0,70	Struktur baik
0,26-0,50	Struktur lemah
≤0,25	Tidak ada struktur substansial yang ditemukan

3. METODE PENELITIAN

Berdasarkan analisis permasalahan, solusi yang diambil menggunakan K-Means clustering seperti pada Gambar 3.



Gambar 3 Framework Penelitian

3.1 Seleksi Fitur

Seleksi fitur merupakan pemilihan atribut yang digunakan sebagai acuan untuk melakukan *clustering*. Penelitian ini mengambil 4 atribut untuk dijadikan bahan penelitian yang terdiri dari GKM, GKNM, IkdK dan IkpK yang diolah menggunakan Microsoft Excel.

3.2 Pre-Processing

Pada tahap *pre-processing* terdapat normalisasi data yang berguna untuk mengubah data dengan rentang nilai yang berbeda menjadi data dengan rentang nilai yang sama, sehingga mudah untuk dibandingkan dan diolah. Normalisasi data dilakukan dengan *min-max normalization* di *python*.

3.3 Penentuan Jumlah Cluster Optimal

Penentuan jumlah *cluster* yang efektif menggunakan metode *Elbow* yang dihasilkan dari perhitungan nilai *Silhouette Coefficient*. Hasil visualisasi grafik *Elbow* yang berada pada titik tertinggi atau mendekati angka satu merupakan jumlah *cluster* yang optimal.

3.4 Binning Data

Binning Data merupakan proses mengelompokkan atau membagi data kontinu ke dalam interval atau kategori diskrit. Tujuan dari *binning* data adalah untuk menyederhanakan data, mengurangi kompleksitas atau membuat representasi yang lebih intuitif.

3.5 Clustering dengan Data Binning

Clustering dilakukan dengan data yang sudah di *binning* dan sudah melewati tahap *pre-processing*. Setelah *bins* dibuat, kelompokkan data ke dalam *bin* yang sesuai berdasarkan nilai atributnya. Kemudian, jalankan *K-Means clustering* untuk mengklaster keempat atributnya dan hasil *cluster* disimpan ke dalam kolom baru. Langkah terakhir yaitu menampilkan hasil *clustering* data yang sudah di *binning*.

3.6 Clustering dengan Data tanpa Binning

Clustering menggunakan algoritma *K-means* dengan nilai *k* yang diperoleh dari metode *Elbow* setelah memperoleh nilai *k* yang optimal. *K-Means* digunakan untuk memilih *centroid cluster* awal menggunakan pengambilan sampel secara acak. Selanjutnya perhitungan jarak setiap *cluster* dengan titik *centroid* dilakukan dengan mengimport *seaborn* yang berfungsi untuk memetakan

setiap data dengan *cluster* terdekat dengan titik pusat *centroid*-nya. Data yang akan di *cluster* dilihat dari seberapa field berpengaruh terhadap field yang lainnya.

3.7 Evaluasi dan Hasil Analisis Cluster

Evaluasi *cluster* menggunakan metode *Silhouette Coefficient* bertujuan untuk melihat kualitas dan kekuatan *cluster* serta mengukur seberapa baik suatu objek ditempatkan dalam suatu *cluster*. Tahapan yang dilakukan dalam perhitungan nilai *Silhouette Coefficient* yaitu dengan menghitung rata-rata jarak dari suatu data dengan data lain yang berada dalam satu *cluster*. Selanjutnya menghitung rata-rata jarak data tersebut dengan semua data di *cluster* lain dan diambil nilai terkecilnya.

Analisis *cluster* merupakan tahap akhir pengklasteran dengan menentukan karakteristik dari setiap *cluster* yang dihasilkan. Pada tahap analisis dijelaskan hasil *cluster* yang terbentuk memiliki karakteristik yang berbeda dengan *cluster* yang lain. Hasil *clustering* yang terbentuk ditampilkan berdasarkan penggunaan data *binning* dan tanpa *binning*. Berdasarkan karakteristik *cluster* yang diperoleh dapat memudahkan bagi pihak terkait untuk mengambil kebijakan dalam mengatasi kemiskinan di masing-masing daerah di Indonesia.

4. HASIL DAN PEMBAHASAN

Berdasarkan hasil dari penelitian, pengelompokkan dengan menggunakan algoritma *K-Means clustering* dan bahasa pemrograman *Python* untuk penelitiannya, yang dibantu dengan tools *Jupyter Notebook*. Dengan pembahasan sebagai berikut:

4.1 Seleksi Fitur

Pada tahap ini dilakukan pemilihan atribut apa saja yang akan digunakan. Berdasarkan data yang diperoleh penelitian ini mengambil atribut Gkm, GKNM, IkdK, dan IkpK.

4.2 Pre-Processing

Pada tahap ini dilakukan penggabungan 4 atribut ke dalam satu file excel tanpa membedakan data perkotaan dan pedesaan sehingga memperoleh hasil 4 file csv siap untuk diolah. Data tersebut terdiri dari data tahun 2019, 2020, dan 2022 semester ke-1 di bulan Maret dan tahun 2021 semester ke-2

di bulan September. *Sample* data setelah melewati tahap *pre-processing* seperti pada Tabel 2.

Tabel 2 *Sample* Data 2019

2019 Maret				
Provinsi	Garis Kemiskinan Makanan (Rupiah/Kapita/Bulan)	Garis Kemiskinan Non-Makanan (Rupiah/Kapita/Bulan)	Indeks Kedalaman Kemiskinan (Persen)	Indeks Keparahan Kemiskinan (Persen)
Aceh	370093.00	116842.00	2.64	0.66
Sumatera Utara	351215.00	114907.00	1.37	0.31
...
Papua Barat	441803.00	131510.00	5.60	1.97
Papua	404631.00	135468.00	7.17	2.60

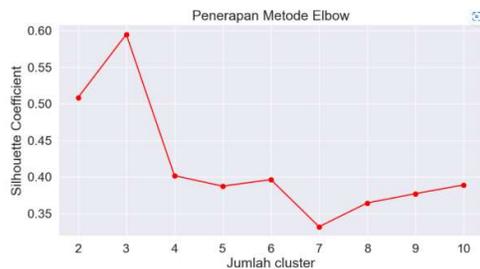
Setelah dilakukan penggabungan data dapat dilanjutkan ke tahap normalisasi data yang menjadi bagian dari tahap *pre-processing*. Normalisasi data dilakukan dengan *min-max normalization* di *python* dengan data tahun 2019-2022. *Sample* hasil normalisasi data seperti pada Tabel 3.

Tabel 3 *Sample* Normalisasi Data 2019

2019 Maret				
Provinsi	Garis Kemiskinan Makanan (Rupiah/Kapita/Bulan)	Garis Kemiskinan Non-Makanan (Rupiah/Kapita/Bulan)	Indeks Kedalaman Kemiskinan (Persen)	Indeks Keparahan Kemiskinan (Persen)
Aceh	0.503507	0.323185	0.320840	0.224
Sumatera Utara	0.427057	0.308715	0.130435	0.084
...
Papua Barat	0.793911	0.432878	0.764618	0.748
Papua	0.643376	0.462477	1.000000	1.000

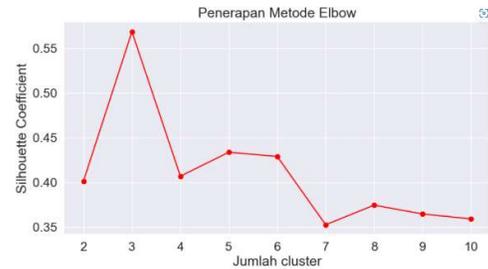
4.3 Penentuan Jumlah *Cluster* Optimal

Penentuan jumlah *cluster* optimal menggunakan metode *Elbow* dari hasil perhitungan nilai *Silhouette Coefficient*. Penggabungan metode tersebut bertujuan untuk memperoleh *cluster* yang baik dan berkualitas. Berikut merupakan hasil visualisasi metode *Elbow*.



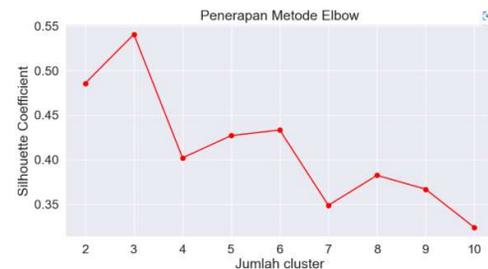
Gambar 4 Visualisasi *Elbow* Data 2019

Berdasarkan hasil visualisasi grafik *elbow* dengan data tahun 2019 memperoleh nilai 0.595.



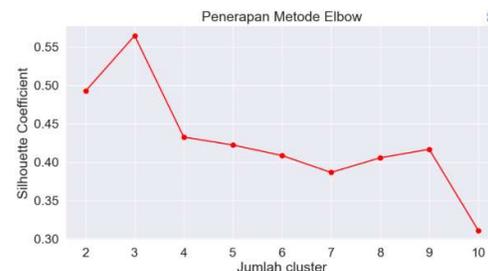
Gambar 5 Visualisasi *Elbow* Data 2020

Berdasarkan hasil visualisasi grafik *elbow* dengan data tahun 2020 memperoleh nilai 0.568.



Gambar 6 Visualisasi *Elbow* Data 2021

Berdasarkan hasil visualisasi grafik *elbow* dengan data tahun 2021 memperoleh nilai 0.540.



Gambar 7 Visualisasi *Elbow* Data 2022

Berdasarkan hasil visualisasi grafik *elbow* dengan data tahun 2022 memperoleh nilai 0.564.

Visualisasi *Elbow* dengan menggunakan data tahun 2019-2022 memperoleh nilai *silhouette* tertinggi dengan rentan 0,540-0,595 yang terletak di *cluster* 3. Oleh karena itu, jumlah *cluster* terbaik adalah $k=3$ sesuai dengan hasil *silhouette coefficient*.

4.4 *Binning* Data

Binning data dilakukan setelah data melalui tahap pra-pemrosesan yaitu normalisasi data. Pada tahap ini, $n_bins=2$ dimana terdapat nilai rendah dengan interval 0,0-0,5 dan nilai

tinggi dengan interval 0,5-1,0. Sample hasil binning data seperti pada Tabel 4.

Tabel 4 Sample Hasil Binning Data 2019

	Garis Kemiskinan Makanan (Rupiah/Kapita/Bulan) Binned	Garis Kemiskinan Non-Makanan (Rupiah/Kapita/Bulan) Binned	Indeks Kedalaman Kemiskinan(Perses) Binned	Indeks Keparahan Kemiskinan(Perses) Binned
0	Bin 2	Bin 1	Bin 1	Bin 1
1	Bin 1	Bin 1	Bin 1	Bin 1
2	Bin 2	Bin 1	Bin 1	Bin 1
...
32	Bin 2	Bin 1	Bin 2	Bin 2
33	Bin 2	Bin 1	Bin 2	Bin 2

4.5 Clustering dengan Data Binning

Berdasarkan hasil *clustering K-Means* menggunakan data *binning* tahun 2019 terdapat 24 provinsi di *cluster_1*, 8 provinsi di *cluster_2*, dan 2 provinsi di *cluster_3*. Dengan nama-nama provinsi berdasarkan kelompoknya yang dapat dilihat pada Tabel 5.

Tabel 5 Hasil Clustering Binning Data 2019

Hasil Clustering Data Tahun 2019		
Cluster	Jumlah	Provinsi
<i>Cluster_1</i>	24	Bali, Sulawesi Selatan, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Barat, Nusa Tenggara Timur, Nusa Tenggara Barat, Sulawesi Tenggara, Banten, Jawa Timur, Di Yogyakarta, Sulawesi Tengah, Jawa Tengah, Gorontalo, Sulawesi Barat, Lampung, Bengkulu, Sumatera Selatan, Jambi, Riau, Maluku Utara, Sumatera Utara, Jawa Barat, dan Sulawesi Utara.
<i>Cluster_2</i>	8	Maluku, Aceh, Kalimantan Timur, Dki Jakarta, Kep. Riau, Kep. Bangka Belitung, Sumatera Barat, dan Kalimantan Utara.
<i>Cluster_3</i>	2	Papua Barat dan Papua.

Berdasarkan hasil *clustering K-Means* menggunakan data *binning* tahun 2020 terdapat 23 provinsi di *cluster_1*, 8 provinsi di *cluster_2*, dan 3 provinsi di *cluster_3*. Dengan nama-nama provinsi berdasarkan kelompoknya yang dapat dilihat pada Tabel 6.

Tabel 6 Hasil Clustering Binning Data 2020

Hasil Clustering Data Tahun 2020		
Cluster	Jumlah	Provinsi
<i>Cluster_1</i>	23	Sulawesi Tenggara, Sulawesi Selatan, Kalimantan Tengah, Sulawesi Barat, Sulawesi Tengah, Sulawesi Utara, Maluku Utara, Kalimantan Selatan, Gorontalo, Kalimantan Barat, Bali, Nusa Tenggara Barat, Banten, Jawa Timur, Di Yogyakarta, Jawa Tengah, Jawa Barat, Lampung, Bengkulu, Sumatera Selatan, Jambi, Sumatera Utara, dan Nusa Tenggara Timur.
<i>Cluster_2</i>	8	Aceh, Kalimantan Timur, Kalimantan Utara, Dki Jakarta, Kep. Riau, Kep. Bangka Belitung, Riau, dan Sumatera Barat.
<i>Cluster_3</i>	3	Papua Barat, Maluku dan Papua.

Berdasarkan hasil *clustering K-Means* menggunakan data *binning* tahun 2021 terdapat 22 provinsi di *cluster_1*, 9 provinsi di *cluster_2*, dan 3 provinsi di *cluster_3*. Dengan nama-nama provinsi berdasarkan kelompoknya yang dapat dilihat pada Tabel 7.

Tabel 7 Hasil Clustering Binning Data 2021

Hasil Clustering Data Tahun 2021		
Cluster	Jumlah	Provinsi
<i>Cluster_1</i>	22	Aceh, Maluku Utara, Sulawesi Barat, Gorontalo, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Utara, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Barat, Nusa Tenggara Barat, Banten, Jawa Timur, Bali, Jawa Tengah, Sumatera Utara, Jambi, Sumatera Selatan, Di Yogyakarta, Lampung, dan Jawa Barat.
<i>Cluster_2</i>	9	Kep. Bangka Belitung, Kalimantan Timur, Kalimantan Utara, Dki Jakarta, Bengkulu, Riau, Sumatera Barat, Maluku, dan Kep. Riau.
<i>Cluster_3</i>	3	Nusa Tenggara Timur, Papua Barat, dan Papua.

Berdasarkan hasil *clustering K-Means* menggunakan data *binning* tahun 2022 terdapat 24 provinsi di *cluster_1*, 8 provinsi di *cluster_2*, dan 2 provinsi di *cluster_3*. Dengan nama-nama provinsi berdasarkan kelompoknya yang dapat dilihat pada Tabel 8.

Tabel 8 Hasil Clustering Binning Data 2022

Hasil Clustering Data Tahun 2022		
Cluster	Jumlah	Provinsi
<i>Cluster_1</i>	24	Aceh, Maluku Utara, Sulawesi Barat, Gorontalo, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Utara, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Barat, Nusa Tenggara Timur, Nusa Tenggara Barat, Banten, Bali, Di Yogyakarta, Sumatera Utara, Jambi, Sumatera Selatan, Bengkulu, Jawa Timur, Lampung, Jawa Barat, dan Jawa Tengah.
<i>Cluster_2</i>	8	Kalimantan Timur, Kalimantan Utara, Kep. Riau, Dki Jakarta, Riau, Sumatera Barat, Maluku, dan Kep. Bangka Belitung.
<i>Cluster_3</i>	2	Papua Barat dan Papua.

4.6 Clustering dengan Data tanpa Binning

Berdasarkan hasil *clustering K-Means* menggunakan data tahun 2019 tanpa dilakukan *binning* terdapat 27 provinsi di *cluster_1*, 5 provinsi di *cluster_2*, dan 2 provinsi di *cluster_3*. Dengan nama-nama provinsi berdasarkan kelompoknya yang dapat dilihat pada Tabel 9.

Tabel 9 Hasil *Clustering* Data 2019

Hasil <i>Clustering</i> Data Tahun 2019		
Cluster	Jumlah	Provinsi
Cluster_1	27	Aceh, Maluku Utara, Maluku, Sulawesi Barat, Gorontalo, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Utara, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Barat, Nusa Tenggara Timur, Nusa Tenggara Barat, Banten, Bali, Di Yogyakarta, Jawa Tengah, Jawa Barat, Sumatera Utara, Sumatera Barat, Riau, Jambi, Jawa Timur, Lampung, Bengkulu, dan Sumatera Selatan.
Cluster_2	5	Kalimantan Timur, Kep.Riau, Dki Jakarta, Kalimantan Utara, dan Kep. Bangka Belitung.
Cluster_3	2	Papua Barat dan Papua.

Berdasarkan hasil *clustering K-Means* menggunakan data tahun 2020 tanpa dilakukan *binning* terdapat 27 provinsi di *cluster_1*, 5 provinsi di *cluster_2*, dan 2 provinsi di *cluster_3*. Dengan nama-nama provinsi berdasarkan kelompoknya yang dapat dilihat pada Tabel 10.

Tabel 10 Hasil *Clustering* Data 2020

Hasil <i>Clustering</i> Data Tahun 2020		
Cluster	Jumlah	Provinsi
Cluster_1	27	Maluku Utara, Maluku, Sulawesi Barat, Gorontalo, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Utara, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Barat, Nusa Tenggara Timur, Nusa Tenggara Barat, Aceh, Jawa Timur, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bali, Jawa Barat, Jawa Tengah, Di Yogyakarta, dan Banten.
Cluster_2	5	Kalimantan Utara, Kalimantan Timur, Kep. Riau, Dki Jakarta, dan Kep. Bangka Belitung.
Cluster_3	2	Papua dan Papua Barat.

Berdasarkan hasil *clustering K-Means* menggunakan data tahun 2021 tanpa dilakukan *binning* terdapat 25 provinsi di *cluster_1*, 5 provinsi di *cluster_2*, dan 4 provinsi di *cluster_3*. Dengan nama-nama provinsi berdasarkan kelompoknya yang dapat dilihat pada Tabel 11.

Tabel 11 Hasil *Clustering* Data 2021

Hasil <i>Clustering</i> Data Tahun 2021		
Cluster	Jumlah	Provinsi
Cluster_1	25	Maluku Utara, Sulawesi Barat, Gorontalo, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Utara, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Barat, Nusa Tenggara Barat, Aceh, Jawa Timur, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bali, Jawa Barat, Jawa Tengah, Di Yogyakarta, dan Banten.
Cluster_2	5	Kalimantan Utara, Kalimantan Timur, Kep. Riau, Dki Jakarta, dan Kep. Bangka Belitung.
Cluster_3	4	Papua, Maluku, Nusa Tenggara Timur, dan Papua Barat.

Berdasarkan hasil *clustering K-Means* menggunakan data tahun 2022 tanpa dilakukan *binning* terdapat 27 provinsi di *cluster_1*, 5 provinsi di *cluster_2*, dan 2 provinsi di *cluster_3*. Dengan nama-nama provinsi berdasarkan kelompoknya yang dapat dilihat pada Tabel 12.

Tabel 12 Hasil *Clustering* Data 2022

Hasil <i>Clustering</i> Data Tahun 2022		
Cluster	Jumlah	Provinsi
Cluster_1	27	Aceh, Maluku Utara, Maluku, Sulawesi Barat, Gorontalo, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Tengah, Sulawesi Utara, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Barat, Nusa Tenggara Timur, Nusa Tenggara Barat, Banten, Bali, Di Yogyakarta, Jawa Tengah, Jawa Barat, Jawa Timur, Lampung, Bengkulu, Sumatera Selatan, Jambi, Riau, Sumatera Barat, dan Sumatera Utara.
Cluster_2	5	Kep. Bangka Belitung, Dki Jakarta, Kep. Riau, Kalimantan Timur, dan Kalimantan Utara.
Cluster_3	2	Papua Barat dan Papua.

4.7 Evaluasi dan Hasil Analisis *Cluster*

Evaluasi *cluster* bertujuan untuk melihat seberapa baik *cluster* yang terbentuk. Berdasarkan hasil nilai *silhouette_score* yang diperoleh data tahun 2019-2022 yang bernilai dengan range 0,54-0,59 maka *clustering* yang terbentuk memiliki struktur yang baik sesuai dengan pengukuran koefisien *silhouette*.

Berdasarkan hasil *clustering K-Means* dengan $k=3$ menggunakan data *binning* dan tanpa *binning* mengalami perbedaan. Hal tersebut disebabkan *binning* mengubah skala data dan menggabungkan nilai-nilai yang berdekatan ke dalam interval yang lebih besar. Terdapat beberapa provinsi yang tanpa *binning* berada di *cluster_1* dan setelah di *binning* berada di *cluster* yang lain seperti pada Tabel 13.

Tabel 13 Perbedaan Hasil *Clustering* Tanpa *Binning* dan *Binning*

2019		
Cluster	Tanpa <i>Binning</i>	<i>Binning</i>
1	Maluku, Aceh, Sumatera Barat	-
2	-	Maluku, Aceh, Sumatera Barat
2020		
Cluster	Tanpa <i>Binning</i>	<i>Binning</i>
1	Aceh, Riau, Sumatera Barat, Maluku	-
2	-	Aceh, Riau, Sumatera Barat
3	-	Maluku
2021		
Cluster	Tanpa <i>Binning</i>	<i>Binning</i>
1	Bengkulu, Riau, Sumatera Barat	-
2	-	Bengkulu, Riau, Sumatera Barat, Maluku
3	Maluku	-
2022		
Cluster	Tanpa <i>Binning</i>	<i>Binning</i>
1	Riau, Sumatera Barat, Maluku	-
2	-	Riau, Sumatera Barat, Maluku

Berdasarkan hasil *clustering* data tanpa *binning* dengan data tahun 2019, 2020, dan 2022 menghasilkan provinsi di setiap *cluster tetap* dan **tidak berubah** dengan *cluster_1* berjumlah 27 provinsi dengan kategori kemiskinan tingkat rendah, *cluster_2* berjumlah 5 provinsi dengan kategori kemiskinan tingkat sedang, dan *cluster_3* berjumlah 2 provinsi dengan kategori kemiskinan tingkat tinggi. Sedangkan, hasil *clustering* data tahun 2019, 2020, dan 2022 menggunakan *binning* data pada data sebelum pandemi, selama pandemi dan sesudah pandemi menghasilkan pergeseran *cluster* yaitu pada provinsi Riau, Maluku, dan Aceh. Pada tahun 2019 Riau berada di *cluster_1* dan Maluku berada di *cluster_2*. Pada tahun 2020 Riau dan Aceh berada di *cluster_2* dan Maluku bergeser di *cluster_3*. Pada tahun 2022 Maluku bergeser di *cluster_2* dan Aceh bergeser di *cluster_1*.

Berdasarkan hasil *clustering* pada puncak pandemi menuju kehidupan kenormalan menggunakan data tanpa *binning* menghasilkan pergeseran *cluster*. Pergeseran terjadi pada provinsi Maluku dan Nusa Tenggara Timur yang sebelumnya pada tahun 2021 berada di *cluster_3* dengan kategori kemiskinan parah pada puncak pandemi dikarenakan pertumbuhan ekonomi di Indonesia yang kurang stabil dan mengalami deflasi. Kemudian, pada kehidupan menuju kenormalan provinsi tersebut kembali bergeser di *cluster_1*

dengan kategori kemiskinan rendah dikarenakan perekonomian di Indonesia menuju pada perekonomian yang kembali stabil. Sedangkan, hasil *clustering* data *binning* menghasilkan pergeseran *cluster*. Pada tahun 2021 provinsi Bengkulu berada di *cluster_2* dan Nusa Tenggara Timur berada di *cluster_3*. Kemudian, pada tahun 2022 kedua provinsi tersebut bergeser di *cluster_1*.

5. KESIMPULAN

1. Sebelum, selama dan sesudah pandemi tanpa *binning* tidak ada pergeseran *cluster*. Setelah di *binning* terjadi pergeseran pada provinsi Riau, Maluku, dan Aceh. Pada tahun 2019 Riau berada di *cluster_1* dan Maluku berada di *cluster_2*. Pada tahun 2020 Riau dan Aceh berada di *cluster_2* dan Maluku bergeser di *cluster_3*. Pada tahun 2022 Maluku bergeser di *cluster_2* dan Aceh bergeser di *cluster_1*.

2. Puncak pandemi menuju kehidupan kenormalan tanpa *binning* terjadi pergeseran *cluster* pada provinsi Maluku dan Nusa Tenggara Timur. Pada tahun 2021 provinsi tersebut berada di *cluster_3* dengan kategori kemiskinan tingkat parah dikarenakan pertumbuhan ekonomi di Indonesia yang kurang stabil dan mengalami deflasi. Kemudian, pada kehidupan menuju kenormalan provinsi tersebut bergeser di *cluster_1* dengan kategori kemiskinan rendah. *Clustering* menggunakan data *binning* juga mengalami pergeseran *cluster*. Pada tahun 2021 provinsi Bengkulu berada di *cluster_2* dan Nusa Tenggara Timur berada di *cluster_3*. Kemudian, pada tahun 2022 kedua provinsi tersebut bergeser di *cluster_1*.

3. Perhitungan kualitas *clustering* yang dihasilkan menggunakan *Silhouette Coefficient* dengan data tahun 2019-2022 memperoleh nilai antara 0,54-0,59 maka *clustering* yang terbentuk memiliki interpretasi yang baik karena mendekati angka 1 sesuai dengan pengukuran koefisien *silhouette*.

6. REFERENSI

- [1] E. Widodo, P. Ermayani, L. N. Laila And A. T. Madani, "Pengelompokan Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan

- Analisis *Hierarchical Agglomerative Clustering*," *Seminar Nasional Official Statistics*, Pp. 557-565, 2021.
- [2] Y. R. Sembiring, Saifullah And R. Winanjaya, "Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma *K-Means*" *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, Vol. 2, No. 2, Pp. 125-132, 2021.
- [3] N. N. F. R, D. S. Anggraeni And U. Enri, "Pengelompokkan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma *K-Means* Dengan *Silhouette Coefficient*," *Jurnal Teknologi Informasi Komunikasi (E-Journal)*, Pp. 29 - 35, 2022.
- [4] BPS, "Badan Pusat Statistik," 2020.[Online].Available: <https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html#subjekviewtab1>. [Accessed Januari 2023].
- [5] Y. R. Sari, A. Sudewa, D. A. Lestari And T. I. Jaya, "Penerapan Algoritma *K-Means* Untuk *Clustering* Data Kemiskinan Provinsi Banten Menggunakan *Rapidminer*," *CESS (Journal Of Computer Engineering System And Science)*, Pp. 192-198, 2020.
- [6] B. M. Metisen And H. L. Sari, "Analisis *Clustering* Menggunakan Metode *K-Means* Dalam Pengelompokkan Penjualan Produk Pada Swalayan Fadhila," *Jurnal Media Infotama*, Vol. 11, Pp. 110-118, 2015.
- [7] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan *K-Means*" *Jiska (Jurnal Informatika Sunan Kalijaga)*, Vol. 2, No. 3, P. 167-174, 2018.
- [8] N. P. E. Merliana, Ernawati And A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik Pada Metode *K-Means Clustering*," *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers Unisbank (SENDI_U)*, 2015.
- [9] F. Marisa, S. S. S. Ahmad, Z. I. M. Yusof, Fachrudin And T. M. A. Aziz, "Segmentation Model Of Customer Lifetime Value In Small And Medium Enterprise (Smes) Using *K-Means Clustering* And *LRFM Model*," *International Journal Of Integrated Engineering*, Pp. 169-180, 2019.
- [10] L. Kaufman And P. J. Rousseeuw,] *Finding Groups In Data: An Introduction To Cluster Analysis*, Canada: Simultraneosly In Canada, 2005.