

**PENGELOMPOKAN PENDUDUK MISKIN
MENGUNAKAN METODE K-MEANS CLUSTERING
(STUDI KASUS : PROVINSI JAWA BARAT)**

**Juliandro Sipayung¹, Markus Patrick Hutajulu², Hilman Juan Carlos Sibarani³,
Erika Immanuela Sianipar⁴**

Institut Teknologi Telkom Purwokerto

E-mail: 20102187@ittelkom-pwt.ac.id¹, 20102019@ittelkom-pwt.ac.id²,
21102269@ittelkom-pwt.ac.id³, 2211111054@ittelkom-pwt.ac.id⁴

Abstract

Clustering analysis is a method aimed at grouping objects with similar characteristics into one cluster, while objects with different characteristics are placed in other clusters. The K-Means algorithm, on the other hand is one of the methods aimed at minimizing a function by reducing variation within one cluster and maximizing variation between clusters. This research addresses poverty issues in West Java Province. One unresolved issue in poverty alleviation is the inconsistency between government survey results and the increasing number of job seekers despite the lack of available job opportunities in West Java province. The challenges faced by the government in addressing poverty include inequality and difficulty finding employment opportunities. Data collection for this research uses the 2022 dataset on population in West Java Province, obtained from Badan Pusat Statistik (BPS). The dataset consists of 9 columns, including variables such as "penduduk_miskin" (poor population) and "jumlah_pencari_kerja" (number of job seekers). Subsequently, an optimal value for K is determined using elbow method and Silhouette Analysis, which results in 2 clusters: Cluster 0 and Cluster 1. This study will utilize the K-Means method to estimate or predict the number of people classified as poor by calculating data obtained from BPS. The objective is to generate clustering of populations categorized as poor so that the government can identify areas that require focused attention within West Java Province. The clustering results indicate that Cluster 0 includes districts/cities where there are high concentrations of poor populations such as Ciamis Regency, Kuningan Regency, Sumedang Regency, West Bandung Regency, Pangandaran Regency, Bogor City, Sukabumi City, Bandung City, Cirebon City, Bekasi City Depok city, Cimahi city, Tasikmalaya city and Banjar city.

Keywords — Clustering Analysis, K-Means Algorithm, Poor Residents.

Abstrak

Analisis merupakan suatu metode yang bertujuan untuk mengelompokkan objek-objek dengan karakteristik yang mirip ke dalam satu klaster, sementara objek dengan karakteristik yang berbeda ditempatkan di klaster lain. Sedangkan Algoritma merupakan salah satu metode yang bertujuan untuk meminimalkan fungsi dengan mengurangi variasi antara data dalam satu klaster dan memaksimalkan variasi data di klaster lainnya. Penelitian ini mengangkat permasalahan kemiskinan di Provinsi Jawa Barat. Salah satu permasalahan yang belum teratasi dalam kemiskinan adalah ketidaksesuaian hasil survei pemerintah dengan jumlah pencari kerja yang kian banyak namun masih kurangnya lapangan pekerjaan yang ada pada provinsi Jawa Barat. Tantangan dalam penanganan kemiskinan oleh pemerintah adalah ketidakmerataan dan sulitnya mencari lapangan pekerjaan. Pengumpulan data pada penelitian ini menggunakan dataset penduduk Provinsi Jawa Barat tahun 2022, dataset ini diambil dari Badan Pusat Statistik () dan terdiri dari 9 kolom dan data yang dipakai adalah penduduk_miskin dan jumlah_pencari_kerja. Selanjutnya dilakukan pencarian nilai K yang optimal menggunakan metode dan dimana menghasilkan 2 klaster yakni klaster 0 dan klaster 1. Penelitian ini akan menggunakan metode

untuk melakukan pendugaan atau prediksi jumlah penduduk yang tergolong miskin dengan menghitung data yang diperoleh dari . Penelitian ini bertujuan untuk menghasilkan klasterisasi penduduk yang tergolong miskin, sehingga pemerintah dapat mengetahui daerah mana saja yang harus menjadi fokus perhatian di Provinsi Jawa Barat. Hasil dari klasterisasi mengindikasikan bahwa kluster 0 dikelompokkan sebagai penduduk miskin pada Kabupaten & Kota Provinsi Jawa Barat diantaranya kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Sumedang, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.

Kata Kunci — Analisis Pengelompokan, Algoritma K-Means, Penduduk Miskin.

1. PENDAHULUAN

Kemiskinan tetap menjadi perhatian utama pemerintah di berbagai negara, termasuk Indonesia. Masalah kemiskinan bahkan memiliki dampak yang signifikan dalam bidang ekonomi dan menjadi tolok ukur keberhasilan suatu negara dari waktu ke waktu, terutama bagi negara-negara yang sedang berkembang seperti Indonesia (Hilmi et al., 2022). Indonesia adalah negara yang memiliki populasi pertambahan penduduk yang sangat tinggi. Setiap tahunnya jumlah penduduk Indonesia mengalami peningkatan (Ferdiana & Saputri, 2023). Masalah kemiskinan menjadi salah satu isu serius yang dihadapi Indonesia, terutama di Provinsi Jawa Barat. Menurut data BPS bulan Maret 2020, persentase penduduk miskin di Provinsi Jawa Barat sebesar 7,88%. Namun, pada bulan September 2020, persentasenya meningkat menjadi 8,43%, mengalami peningkatan sebesar 0,55% dalam kurun waktu satu tahun (Dani & Ni' matuzzahroh, 2021).

Kemiskinan menyebabkan banyak anak tidak dapat mengakses pendidikan berkualitas, mengalami kesulitan dalam membiayai kesehatan, memiliki tabungan yang minim dan kurangnya investasi, serta sulit mendapatkan akses ke layanan publik. Selain itu, kemiskinan juga berdampak pada terbatasnya lapangan pekerjaan, ketidakjaminan sosial dan perlindungan keluarga. Fenomena urbanisasi semakin meningkat di perkotaan sebagai akibat dari kemiskinan. Yang lebih serius lagi adalah dampak kemiskinan terhadap jutaan orang yang hanya bisa memenuhi kebutuhan pangan, sandang, dan tempat tinggal secara terbatas. Masalah kemiskinan merupakan salah satu fokus kebijakan pembangunan di setiap negara dengan tujuan untuk mengurangi kesenjangan pendapatan (Amali & Devita, 2022).

Salah satu faktor yang menyebabkan kemiskinan adalah tingginya tingkat pengangguran yang membuat sulitnya mencari pekerjaan (Amali & Devita, 2022).

Pencarian pekerjaan merupakan proses yang menghubungkan individu dengan pekerjaan yang sesuai dengan kualifikasinya (Pa et al., 2022). Dengan adanya populasi tenaga kerja yang signifikan, pergerakan migrasi yang terus berlanjut, dan masih berlangsungnya dampak krisis ekonomi saat ini, masalah ketenagakerjaan menjadi semakin besar dan rumit (Dewita Hia, 2015). Jika semua pekerja dan jenis pekerjaan memiliki kesamaan, sehingga setiap pekerja cocok dengan setiap jenis pekerjaan, maka tidak akan ada masalah dalam mencari pekerjaan. Dalam konteks lembaga pemerintah, data tentang pencari kerja dapat diperoleh dari data yang sudah ada, seperti jumlah pendaftar berdasarkan usia dan pendidikan. Setiap pendaftar memiliki kriteria yang berbeda termasuk pengalaman kerja yang bervariasi. Proses ini terus berulang dan akumulasi data tersebut bisa menghambat pencarian informasi terkait data tersebut (Pa et al., 2022).

Maka dari itu, penulis berupaya memanfaatkan koleksi data yang sudah ada dan menggali informasi yang relevan untuk menentukan jumlah orang yang sedang mencari pekerjaan. Pendekatan ini bertujuan untuk mengatasi kendala dalam mendapatkan informasi penting dari kumpulan data yang dikenal sebagai data mining. Selanjutnya

dilakukan analisis menggunakan metode pengelompokan untuk mengidentifikasi Kabupaten dan Kota di Provinsi Jawa Barat yang memiliki tingkat kemiskinan yang tinggi (Pa et al., 2022).

Clustering adalah proses memisahkan informasi dalam sebuah himpunan menjadi kelompok-kelompok yang memiliki kesamaan informasi di dalam kelompok tersebut lebih besar. Dari pada kesamaan informasinya dengan kelompok lain (Rofiqo et al., 2018). Kemampuan k-ling dapat digunakan untuk mengenali pola-pola dalam informasi yang bermanfaat dalam berbagai aplikasi, seperti pengaturan foto, klasifikasi data, dan penentuan pola. Clustering merupakan metode klasifikasi tanpa pengawasan yang membagi sekelompok objek data menjadi beberapa kelas. Umumnya, pengelompokan (clustering) sering digunakan dalam text mining dan analisis web. Metode pengelompokan (clustering) beroperasi dengan cara yang relatif sederhana, yaitu memisahkan data ke dalam kelompok-kelompok yang memiliki kesamaan antara objek-objeknya (Sari et al., 2020). Tujuan dari analisis pengelompokan (clustering) adalah untuk memisahkan objek-objek dengan karakteristik yang mirip ke dalam satu kelompok, sementara objek-objek dengan karakteristik yang berbeda ditempatkan di kelompok lainnya. Hal ini dilakukan dengan maksud untuk memaksimalkan kesamaan antar objek dalam satu klaster dan pada saat yang sama meminimalkan kesamaan antar klaster (Nuryani & Darwis, 2021).

Algoritma merupakan algoritma yang digunakan untuk mengelompokkan data dengan cara membagi data ke dalam beberapa kelompok, di mana data-data yang memiliki kesamaan ditempatkan dalam satu kelompok. Data yang berbeda akan dikelompokkan ke dalam kelompok yang berbeda pula. Tujuan dari algoritma K-Means adalah untuk meminimalkan fungsi dengan mengurangi variasi antara data dalam satu klaster dan pada saat yang sama memaksimalkan variasi antara data di klaster lainnya (Hardiani, 2022). Data yang digunakan dalam penelitian ini berdasarkan (Badan Pusat Statistik) Provinsi Jawa Barat (Ferdiana & Saputri, 2023).

2. METODE

A. Identifikasi Masalah

Langkah pertama yang dilakukan dalam penelitian ini adalah mengidentifikasi permasalahan sebagai dasar untuk memahami dan menentukan solusi serta metode penyelesaiannya. Berdasarkan penjelasan dalam latar belakang, penulis memilih untuk mengangkat permasalahan kemiskinan di provinsi Jawa Barat.

Salah satu permasalahan yang belum teratasi dalam kemiskinan adalah ketidaksesuaian hasil survei pemerintah dengan jumlah pencari kerja yang kian banyak namun masih kurangnya lapangan pekerjaan yang ada pada provinsi Jawa Barat. Tantangan dalam penanganan kemiskinan oleh pemerintah adalah ketidakmerataan dan sulitnya mencari lapangan pekerjaan. Hal ini disebabkan oleh kurangnya validasi data yang sering diabaikan, sehingga data yang dihasilkan kurang akurat. Penulis akan menggunakan metode untuk melakukan pendugaan atau prediksi jumlah penduduk yang tergolong miskin dengan menghitung data yang diperoleh dari BPS Provinsi Jawa Barat.

Penelitian ini bertujuan untuk menghasilkan klasterisasi penduduk yang tergolong miskin, sehingga pemerintah dapat mengetahui jumlah penduduk miskin di Provinsi Jawa Barat. Hasil pengelompokan ini diharapkan dapat menjadi acuan fokus dalam upaya mengurangi kemiskinan dan memperbanyak lapangan pekerjaan baru di Provinsi Jawa Barat.

B. Studi Literatur

Pada tahap ini, akan dilakukan studi mendalam terkait konsep dasar dari algoritma-algoritma yang akan digunakan dalam penelitian ini. Sumber literatur yang akan digunakan mencakup jurnal dan buku yang relevan dengan topik penelitian, yaitu kemiskinan dan metode algoritma K-Means.

Berikut adalah beberapa penelitian yang sudah dilakukan menggunakan metode algoritma K-Means; Penelitian yang dilakukan Aditia Yudhistira, Rio Andika dengan judul “Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering” Penelitian ini membahas tentang penerapan metode data mining untuk membentuk cluster siswa berdasarkan nilai akademik, sikap, dan disiplin. Tujuan penelitian ini adalah mendapatkan cluster siswa yang dapat digunakan sebagai acuan dalam meningkatkan nilai siswa di masa depan. Metode K-Means Clustering digunakan untuk mengelompokkan data nilai siswa menjadi tiga cluster. Hasil pengujian menggunakan elbow method menunjukkan bahwa tiga merupakan jumlah yang optimal. Selain itu, hasil pengujian menggunakan juga menunjukkan bahwa tiga cluster memberikan kualitas yang lebih baik dibandingkan dengan jumlah cluster lainnya (Yudhistira & Andika, 2023). Selanjutnya penelitian yang dilakukan Muhammad Reza Fahlevi, Dini Ridha Dwiki Putri, Elvin Syahrin dengan judul “Analisis Pengelompokan Data Pelelangan Barang Dengan Metode K-Means Clustering” Penelitian ini membahas tentang pentingnya data pelelangan dalam perusahaan jasa penyalur kredit. Banyaknya jumlah data pelelangan menyebabkan masalah pengelolaan dan sulitnya mengelompokkan data sesuai dengan jumlah uang yang dipinjam. Aplikasi ini menggunakan untuk membantu pegawai mendapatkan informasi dan mengelompokkan data lelang berdasarkan kriteria tertentu. Pengujian aplikasi menggunakan RapidMiner menunjukkan hasil berdasarkan jenis pinjaman: 17 nasabah dengan pinjaman kecil, 5 nasabah dengan pinjaman sedang, dan 2 nasabah dengan pinjaman besar (Fahlevi et al., 2023).

Selanjutnya adalah penelitian yang dilakukan Candra Adi Rahmat, Hilda Permatasari, Errisya Rasywir, Yovi Pratama dengan judul “Penerapan Untuk Clustering Kondisi Gizi Balita Pada Posyandu” Dalam penelitian ini, dibahas tentang permasalahan kurang gizi pada anak yang menjadi isu utama kesehatan di negara-negara berkembang termasuk Indonesia. Berdasarkan data nasional, ditemukan bahwa 36,8 balita mengalami stunting atau pertumbuhan terhambat. Untuk memahami lebih lanjut mengenai kondisi gizi balita tersebut, digunakan analisis data mining dengan teknik pengelompokan (). Algoritma K-Means digunakan untuk membagi 20 data balita dalam penelitian ini. Hasilnya menunjukkan bahwa terdapat 4 balita dengan status gizi buruk, 7 balita dengan status gizi baik, dan 9 balita yang mengalami obesitas setelah dilakukan pengklusteran secara acak menggunakan tiga pusat kluster (Rahmat et al., 2023).

C. Pengumpulan Data

Data yang diperlukan berasal dari Badan Pusat Statistik (BPS) tahun 2022 untuk Provinsi Jawa Barat yang mencakup id, kode_provinsi, kode_kabupaten_kota, nama_kabupaten_kota, penduduk_miskin, jumlah_pencari_kerja, tahun, satuan.

D. Data

Data yang telah dikumpulkan kemudian mengalami proses pra-pemrosesan sebelum masuk ke tahap klasifikasi. Pada tahap pra-pemrosesan ini, data mentah yang telah berhasil diambil sebelumnya akan diolah terlebih dahulu. Hal ini dilakukan karena data yang diterima masih berupa data acak dan perlu dikelompokkan berdasarkan kebutuhan pengolahan data selanjutnya. Pada tahap ini, dilakukan penentuan kolom yang akan menjadi acuan dalam klusterisasi. Dalam penelitian ini, kolom "Penduduk_Miskin" dan "Jumlah_Pencari_Kerja" digunakan sebagai acuan untuk menentukan kluster tingkat kemiskinan Provinsi Jawa Barat.

Proses pra-pemrosesan memiliki fungsi untuk menghapus atau menghilangkan data yang tidak relevan, serta secara keseluruhan memodifikasi data agar lebih mudah diproses oleh sistem. Dengan kata lain, tahap ini adalah tahap awal untuk menyamakan format data yang telah dikumpulkan.

E. Implementasi

Setelah melalui tahapan awal identifikasi masalah, studi literatur, pengumpulan data, dan preprocessing, langkah selanjutnya adalah menerapkan metode K-Means pada data yang telah melewati tahapan-tahapan sebelumnya.

F. Pengujian dan Hasil

Pengujian hipotesis dalam penelitian ini, dilakukan proses pengklasteran data menggunakan dua kelompok yang akan diberi label "0" dan "1". Label tersebut akan mengindikasikan rentang nilai dari terendah hingga tertinggi. Hasil klasterisasi data akan digabungkan dengan variabel-variabel kemiskinan, dan hasil klasifikasi klaster akan ditentukan berdasarkan jenis label yang diberikan.

Validitas hasil klasterisasi akan dievaluasi menggunakan silhouetten coefficient untuk mengukur kualitas klaster (Alfiansyah et al., 2022).

3. HASIL DAN PEMBAHASAN

Dataset

Dataset penduduk Jawa Barat pada tahun 2022 yang digunakan untuk pengklusteran terdiri dari 27 data dengan format CSV. Dataset ini diambil dari Badan Pusat Statistik (BPS) tahun 2022 dan terdiri dari 9 kolom, yaitu ID, kode_provinsi, nama_provinsi, kode_kabupaten_kota, nama_kabupaten_kota, penduduk_miskin, jumlah_pencari_kerja, tahun, satuan.

ID	Kode Provinsi	Nama Provinsi	Kode Kabupaten Kota	Nama Kabupaten Kota	Penduduk Miskin	Jumlah Pencari Kerja	Tahun	Satuan
0	32	JAWA BARAT	3201	KABUPATEN BOJONEgara	47810	21245	2022	ORANG
1	32	JAWA BARAT	3202	KABUPATEN SUKATANI	192200	26383	2022	ORANG
2	32	JAWA BARAT	3203	KABUPATEN CIANJUR	248800	26870	2022	ORANG
3	32	JAWA BARAT	3204	KABUPATEN BANDUNG	233600	11317	2022	ORANG
4	32	JAWA BARAT	3205	KABUPATEN GARUT	278100	22882	2022	ORANG
5	32	JAWA BARAT	3206	KABUPATEN TASIKMALAYA	181100	11337	2022	ORANG
6	32	JAWA BARAT	3207	KABUPATEN CIAMIS	34000	5592	2022	ORANG
7	32	JAWA BARAT	3208	KABUPATEN KURUMAN	140300	12824	2022	ORANG
8	32	JAWA BARAT	3209	KABUPATEN CIREBON	288100	43425	2022	ORANG
9	32	JAWA BARAT	3210	KABUPATEN MAJALINGRA	117100	18446	2022	ORANG
10	32	JAWA BARAT	3211	KABUPATEN SUMEDANG	120100	18453	2022	ORANG
11	32	JAWA BARAT	3212	KABUPATEN INDRAMAYU	225000	16189	2022	ORANG
12	32	JAWA BARAT	3213	KABUPATEN SUBANG	155300	27187	2022	ORANG
13	32	JAWA BARAT	3214	KABUPATEN PURBOKARTI	83000	26888	2022	ORANG
14	32	JAWA BARAT	3215	KABUPATEN KARAWANG	188800	27483	2022	ORANG
15	32	JAWA BARAT	3216	KABUPATEN DEKASI	261100	26163	2022	ORANG
16	32	JAWA BARAT	3217	KABUPATEN BANDUNG BARAT	182700	32355	2022	ORANG
17	32	JAWA BARAT	3218	KABUPATEN PADJARAN	87000	2782	2022	ORANG
18	32	JAWA BARAT	3219	KOTA BOGOR	78200	3233	2022	ORANG
19	32	JAWA BARAT	3220	KOTA SUKATANI	35500	3433	2022	ORANG
20	32	JAWA BARAT	3221	KOTA BANDUNG	193800	5555	2022	ORANG
21	32	JAWA BARAT	3222	KOTA CIREBON	87600	2476	2022	ORANG
22	32	JAWA BARAT	3223	KOTA DEKASI	137100	12313	2022	ORANG
23	32	JAWA BARAT	3224	KOTA CIPOK	44100	5763	2022	ORANG
24	32	JAWA BARAT	3225	KOTA CIMAHI	31200	2175	2022	ORANG
25	32	JAWA BARAT	3226	KOTA LAMPALAYA	67700	4892	2022	ORANG
26	32	JAWA BARAT	3227	KOTA BULAR	17700	1517	2022	ORANG

Gambar 1. Dataset awal

Preprocessing

Data yang akan digunakan sebagai acuan dalam pengklusteran adalah data "Penduduk_Miskin" dan "Jumlah_Pencari_Kerja". Data tersebut akan dipecah dan dibentuk menjadi array, kemudian disimpan dalam variabel dataset.

```
array([[474700, 21248],
       [186300, 20383],
       [246800, 30070],
       [258600, 11317],
       [276700, 22682],
       [194100, 14337],
       [ 94000,  9902],
       [140300, 13824],
       [266100, 43428],
       [147100, 18996],
       [120100, 10450],
       [225000, 16189],
       [155300, 27167],
       [ 83400, 20008],
       [199900, 27400],
       [201100, 30163],
       [183700,  3355],
       [ 37900,  2762],
       [ 79200,  3220],
       [ 26600,  3930],
       [109800,  9855],
       [ 31500,  2475],
       [137400, 12318],
       [ 64400,  5263],
       [ 31200,  2175],
       [ 87100,  4652],
       [ 12700,  1812]], dtype=int64)
```

Gambar 2. Pembentukan Array Data

Berikutnya yakni melakukan normalisasi data sebelum melakukan perhitungan menggunakan metode klusterisasi. Normalisasi data dilakukan dengan menggunakan fungsi "MinMaxScaler()" yang terdapat pada library "sklearn.preprocessing". Hasil Normalisasi data dapat dilihat pada Gambar 3.

```
In [5]: from sklearn.preprocessing import MinMaxScaler

In [6]: from sklearn import preprocessing

In [7]: min_max_scaler = preprocessing.MinMaxScaler ( feature_range = (0,1) )

In [8]: col = dataset.columns

In [9]: result = min_max_scaler.fit_transform(dataset)
min_max_scaler_df = pd.DataFrame(result, columns=col)

In [10]: print(min_max_scaler_df)
```

	jumlah_pencari_kerja	penduduk_miskin
0	0.467012	1.000000
1	0.448247	0.575758
2	0.679018	0.586710
3	0.228908	0.532251
4	0.501400	0.571420
5	0.300966	0.392641
6	0.194396	0.175976
7	0.288630	0.276100
8	1.000000	0.548485
9	0.412910	0.290909
10	0.207564	0.232450
11	0.545466	0.459526
12	0.689281	0.586658
13	0.437256	0.150850
14	0.614880	0.485195
15	0.681252	0.407792
16	0.037077	0.370130
17	0.022820	0.054545
18	0.033800	0.163939
19	0.050894	0.050887
20	0.195287	0.210175
21	0.015051	0.040693
22	0.252451	0.260913
23	0.082925	0.111905
24	0.086725	0.040845
25	0.068243	0.161050
26	0.000000	0.000000

Gambar 3. Normalisasi Data

Proses Clustering

Proses pengklusteran dalam Jupyter Notebook menggunakan bahasa pemrograman Python dimulai dengan mencari nilai K yang optimal menggunakan metode *Elbow* dengan *random state* : 42.

```

from sklearn.cluster import KMeans

wcss = []

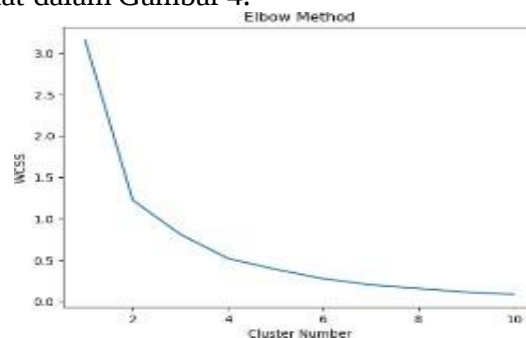
for i in range(1,11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)

plt.plot(range(1,11), wcss)
plt.title('Elbow Method')
plt.xlabel('Cluster Number')
plt.ylabel('WCSS')
plt.show()

```

Gambar 4 Proses mencari nilai K

K nilai optimal ditentukan ketika grafik percobaan pada metode Elbow menunjukkan penurunan yang signifikan dan hampir membentuk sudut tajam. Grafik percobaan inidapat dilihat dalam Gambar 4.



Gambar 5. Proses mencari nilai K

Nilai K optimal yang diperoleh dari metode *elbow curve* seringkali bersifat "ambigu" atau belum pasti akan menghasilkan jumlah *cluster* (K) yang optimal maka dari itu selanjutnya melakukan pencarian Nilai K optimal menggunakan *Silhouette Analysis* yang dapat dilihat pada Gambar 5.

```

range_n_clusters = [2, 3, 4, 5, 6, 7, 8, 9, 10]
for num_clusters in range_n_clusters:

    # Initialize kmeans
    kmeans = KMeans(n_clusters=num_clusters, max_iter=50)
    kmeans.fit(X)
    cluster_labels = kmeans.labels_

    # Silhouette Score
    silhouette_avg = silhouette_score(X, cluster_labels)
    print("For n_clusters={0}, the silhouette score is {1}".format(num_clusters, silhouette_avg))

For n_clusters=2, the silhouette score is 0.5139543271031752
For n_clusters=3, the silhouette score is 0.4318920219098065
For n_clusters=4, the silhouette score is 0.45603629547126004
For n_clusters=5, the silhouette score is 0.3940494800968836

For n_clusters=6, the silhouette score is 0.38430300247455157
For n_clusters=7, the silhouette score is 0.39327125920819267
For n_clusters=8, the silhouette score is 0.42617366057885514
For n_clusters=9, the silhouette score is 0.41956607120367114
For n_clusters=10, the silhouette score is 0.4299459254749543

```

Gambar 6. Silhouette Analysis

Berdasarkan hasil analisis, dapat dilihat bahwa dalam kasus ini, nilai tertinggi untuk *Silhouette score* terjadi ketika *n_clusters* = 2 dengan nilai 0.514, K 2 dipilih karena nilai *silhouette coefficient* mendekati nilai 1. Oleh karena itu, jumlah cluster yang optimal adalah 2 yang menunjukkan bahwa pengelompokan menjadi lebih baik ketika data dibagi menjadi dua kelompok.

```

kmeans = KMeans(n_clusters=2, init='k-means++', random_state=42)
y_means = kmeans.fit_predict(X)
y_means

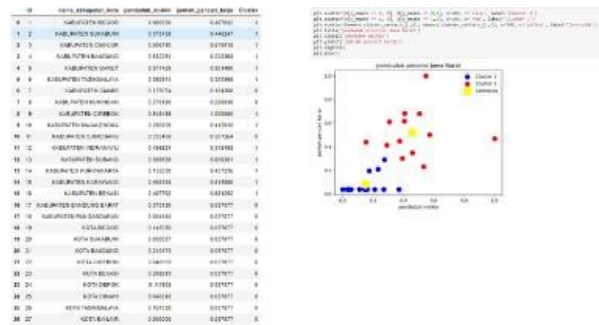
array([1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0])

```

Gambar 7. Klasterisasi Data

Sebagian data yang telah mengalami proses klasterisasi dengan menggunakan metode *K-Means* dengan nilai *k* = 2. Dalam Gambar 8, terdapat kolom "*Cluster*" yang menunjukkan klaster mana setiap data termasuk, yaitu klaster 0 atau klaster 1 lalu

Kedu ini diberi warna merah dan biru untuk warna kuning yaitu centroid. (titik pusat klusterisasi).



Gambar 8. Klasterisasi Data

Berdasarkan **Gambar 8**, semua data yang termasuk dalam kluster "0" adalah kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Sumedang, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar. Kemudian semua data yang termasuk dalam kluster "1" adalah Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi.

Setelah proses optimasi dan penentuan titik centroid pada hasil klasterisasi, berikut adalah titik Centroid yang dihasilkan.

- a. a. Cluster 1 : 0.15122136 0.07846043
- b. b. Cluster 2 : 0.15122136 0.07846043

```
clusters = kmeans.cluster_centers_

print(clusters)

[[0.15122136 0.07846043]
 [0.45787554 0.517242  ]]
```

Gambar 9. Titik Centroids

KESIMPULAN

Berdasarkan hasil penghitungan menggunakan metode K-Means, dapat disimpulkan sebagai berikut:

1. Pengelompokan penduduk miskin pada Kabupaten & Kota Provinsi Jawa Barat pada tahun 2022 berdasarkan variabel penduduk_miskin dan jumlah_pencari_kerja.
2. Dataset penduduk Provinsi Jawa Barat yang digunakan untuk pengklusteran terdiri dari 27 data dengan format CSV Dataset ini diambil dari Badan Pusat Statistik (BPS) dan terdiri dari 9 kolom yang menghasilkan 2 kluster yakni kluster 0 dan kluster 1.
3. Dengan menggunakan metode K-Means bisa digunakan untuk mengelompokkan penduduk miskin Provinsi Jawa Barat.
4. Hasil dari klasterisasi mengindikasikan bahwa kluster 0 dikelompokkan sebagai penduduk miskin pada Kabupaten & Kota Provinsi Jawa Barat diantaranya kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Sumedang, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.

5. Dengan hasil dari pengklusteran ini diharapkan pemerintah lebih memperhatikan wilayah pada provinsi Jawa Barat terutama pada Kabupaten & Kota Provinsi yang masuk kedalam klaster 0, agar kedepannya banyak lapangan pekerjaan baru yang nantinya bisa memperkecil tingkat kemiskinan pada Provinsi Jawa Barat.

REFERENCES

- Alfiansyah, D. N., Nastiti, V. R. S., & Hayatin, N. (2022). Penerapan Metode K-Means pada Data Penduduk Miskin Per Kecamatan Kabupaten Blitar. *Jurnal Repositor*, (1), 49– 58. <https://doi.org/10.22219/repositor.v4i1.1416>
- Amali, M., & Devita, A. (2022). Faktor-Faktor yang Mempengaruhi Kemiskinan di Provinsi Jambi. *J-MAS (Jurnal Manajemen Dan Sains)*, 7(2) 1247. <https://doi.org/10.33087/jmas.v7i2.537>
- Dani, A. T. R., & Ni' matuzzahroh, L. (2021). Pemodelan Persentase Penduduk Miskin Kabupaten/Kota di Provinsi Jawa Barat dengan Pendekatan Regresi Nonparametrik Spline Truncated. *J Statistika : Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 14(1), 24 – 29. <https://doi.org/10.36456/jstat.vol14.no1.a384>
- Dewita Hia, Y. (2015). Strategi Dan Kebijakan Pemerintah Dalam Menanggulangi Pengangguran. *Econmica* 1(2), 208 – 213. <https://doi.org/10.22202/economica.2013.v1.i2.121>
- Fahlevi, M. R., Ridha, D., Putri, D., & Syahrin, E. (2023). Analisis Pengelompokan Data Pelelangan Barang Dengan Metode Clustering. *Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika)*, 8(1), 53 – 61. <http://tunasbangsa.ac.id/ejurnal/index.php/jurasik/article/view/541>
- Ferdiana, K., & Saputri, V. A. (2023). Analisis Clustering Kabupaten / Kota di Jawa Timur Berdasarkan Kondisi Sosial Ekonomi dengan Metode Clustering Analysis of Cities in East Java Based on Socio – Economic Conditions with the K-Means Method. *JIKOSTIK – Jurnal Ilmiah Komputai Dan Statistika e – ISSN : 2087-3657*, Volume 2, Nomor 2, Februari 2023, 2, 1 – 9.
- Hardiani, T. (2022). Analisis Clustering Kasus Covid 19 d Indonesia Menggunakan Algoritma K-Means . *Jurnal Nasonal Pedidikan Teknik Informatika (JANAPATI)*, 11(2), 156 – 165 <https://doi.org/10.23887/janapati.v11i2.45376>
- Hilmi, Marumu, M. N. H. D., Ramlawati, & Peuru, C. D. (2022). Pengaruh Jumlah Penduduk Dan Pengangguran Terhadap Tingkat Kemiskinan Di Kabupaten Tolitoli. *Growth : Jurnal Ilmiah Ekonomi Pembangunan*, 1(1), 20– 27.
- Nuryani, I., & Darwis, D. (2021). Analisis pada Pengguna Brand HP Menggunakan Metode K-Means. *Prosiding Seminar Nasional Ilmu Komputer*, 1 (1), 190 – 211
- Pa, P., Pardede, A. M. H., & Suci, R. (2022). Pengelompokan Data Pencari Kerja Terdaftar Berdasarkan Umur Dan Pendidikan Menggunakan Metode K-Means Clustering Di Dinas Tenaga Kerja Dan Perindustrian Perdagangan Kota Binjai. *Jurnal Informatika Kaputama*, (3), 799–6808.
- Rahmat, C. A., Permatasari, H., & ... (2023). Penerapan Untuk Kondisi Gizi Balita Pada Posyandu. *Jurnal Media ...*, 7 , 207– 213. <https://doi.org/10.30865/mib.v7i1.5142>
- Sari, Y. R., Sudewa, A., Lestari, D. A., & Jaya, T. I. (2020). Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer. *CESS (Journal of Computer Engineering, System and Science)*, 5(2),192. <https://doi.org/10.24114/cess.v5i2.18519>
- Yudhistira, A., & Andika, R. (2023). Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *Journal of Artificial Intelligence and Technology Information*,(1), 20 – 28. <https://doi.org/10.58602/jaiti.v1i1.22>.