IMPLEMENTASI K-MEANS CLUSTERING UNTUK PENGELOMPOKAN TINGKAT KEMISKINAN DI PULAU JAWA MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON

Inggit Agustina, Raditya Danar Dana
Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon
Jalan Perjuangan No. 10B Karyamulya Kec. Kesambi Kota Cirebon, Jawa barat 45131
Inggitagustina2@gmail.com

ABSTRAK

Kemiskinan adalah masalah ekonomi sosial yang sudah menyebar di seluruh wilayah di Indonesia, termasuk di Pulau Jawa yang merupakan pulau terpadat di Indonesia dengan kepadatan penduduk tercatat sebesar 1.015,9 jiwa/km2 dan menjadi pulau di Indonesia dengan jumlah penduduk miskin terbanyak , yaitu 13,62 juta orang. Bedasarkan data yang diperoleh dari Badan Pusat Statistik (BPS). Terlihat sebaran kemiskinan di seluruh wilayah Indonesia, tetapi, data yang disajikan oleh BPS hanya memaparkan presentase bedasarkan provinsi. Oleh karena itu, penelitian ini dilakukan dengan tujuan mengelompokkan kabupaten atau kota di Pulau Jawa berdasarkan tingkat kemiskinannya. penelitian ini menggunakan data sekunder data dan informasi kemiskinan tahun 2022 menggunakan metode *clustering* dengan algoritma *K-Means*. Metode validasi cluster menggunakan Davies Bouldien-Index dengan nilai 0,9315 dengan bantuan tools anaconda dan bahasa pemrograman python. Hasil Penelitin ini menunjukkan kabupaten/kota di Pulau Jawa dapat di kelompokkan menjadi dua yaitu, (cluster 0) kabupaten/kota dengan tingkat kemiskinan rendah terdiri dari 86 kabupaten/kota dengan rata-rata tingkat kemiskinan, yaitu 8,86% dan rata-rata kedalaman kemiskinan(P1) 1,31%. Cluster (1) kabupaten/kota dengan tingkat kemiskinan tinggi terdiri dari 26 kabupaten/kota dengan rata-rata tingkat kemiskinan, yaitu 10,63% dan rata-rata kedalaman kemiskinan(P1), yaitu 1,90%. Kabupaten Bogor dengan tingkat kemiskinan paling tinggi dan Kota Mojokerto dengan tingkat kemiskinan paling rendah.

Kata kunci : Kemiskinan, Data Mining, *K-Means, Clustering, Davies Bouldien-Index*

1. PENDAHULUAN

Indonesia, sebagai negara maritim terluas di kawasan ASEAN, memiliki wilayah seluas 1,9 juta kilometer persegi dan terdiri dari 17.000 pulau. Populasinya mencapai sekitar 278,70 juta jiwa pada tahun 2023. Kemiskinan adalah tantangan sosial yang dihadapi oleh Indonesia, Berdasarkan Undang-Undang No. 24 Tahun 2004, Kemiskinan menjadi isu sosial yang dihadapi oleh negara Indonesia. Kemiskinan dapat diartikan sebagai situasi di mana individu atau kelompok yang tidak mamput memenuhi kebutuhan dasar mereka untuk hidup dengan nyaman. Kebutuhan dasar tersebut mencakup aspek-aspek seperti makanan, kesehatan, pendidikan, pekerjaan, tempat tinggal, pasokan air, pertanahan, sumber daya alam, lingkungan hidup, keamanan, dan partisipasi dalam kehidupan bermasyarakat.[1]

ada berbagai faktor yang menjadi penyebab kemiskinan kemiskinan, tetapi secara umum dapat diklasifikasikan menjadi dua kelompok, yaitu faktor ekonomi dan faktor sosial. Faktor ekonomi yang penyebab kemiskinan melibatkan kurangnya akses terhadap produksi. Akses terhadap pekerjaan yang rendah dapat disebabkan oleh rendahnya keterampilan, keterbatasan lapangan kerja, dan diskriminasi. Akses terhadap produksi yang rendah dapat disebabkan oleh rendahnya modal usaha, lemahnya akses terhadap pasar, dan terbatasnya penguasaan teknologi. Faktor sosial yang menyebabkan kemiskinan antara lain rendahnya pendidikan, rendahnya kesehatan, dan diskriminasi. Pendidikan yang rendah dapat menyebabkan seseorang tidak memiliki keterampilan yang dibutuhkan untuk bekerja. Kesehatan yang rendah dapat menyebabkan seseorang tidak produktif dan tidak dapat bekerja. Diskriminasi dapat menyebabkan seseorang tidak mendapatkan kesempatan yang sama untuk mengakses pekerjaan dan sumber daya lainnya. Pendapatan seseorang akan menentukan besarnya pengeluaran konsumsi. Semakin tinggi penghasilan individu, semakin besar pula belanjanya untuk konsumsi. Hal ini karena seseorang akan memiliki lebih banyak uang untuk memenuhi kebutuhan dasar dan kebutuhan lainnya.[2]

Kemiskinan permasalahan yang banyak dihadapi negara, terutama negara yang masih berkembang dan tertinggal. Kemiskinan merupakan masalah yang rumit, dipengaruhi oleh berbagai faktor, tidak hanya aspek ekonomi, melainkan juga dimensi politik, sosial, budaya, dan struktur sosial lainnya. Dalam konteks ini, kemiskinan tidak hanya terbatas pada ekonomi, tetapi juga melibatkan aspek-aspek lainnya. Kemiskinan dijelaskan sebagai kondisi di mana individu atau keluarga menghadapi kesulitan memenuhi kebutuhan dasar mereka. Selain itu, lingkungan di sekitar penduduk miskin tidak memberikan peluang yang cukup untuk meningkatkan kesejahteraan atau keluar dari keadaan rentan. Hal ini menunjukkan bahwa kemiskinan tidak hanya masalah materi, tetapi juga terkait dengan faktor-faktor lingkungan dan sistemik

yang dapat mempengaruhi kondisi sosio-ekonomi secara keseluruhan.[3]

Penelitian ini bedasarkan pada tinjauan literatur terhadap penelitian serupa sebelumnya k dijadikan salah satu acuan sebagai opsi pengambilan keputusan. Nova Novitasari telah melakukan penelitian yang dipublikasikan, dalam jurnal informatika terpadu, pada tahun 2020 dengan judul "Penerapan Algoritma K-Means Untuk Clustering Data Jumlah Penduduk Miskin Berdasarkan Kota/Kabupaten jumlah penduduk miskin di KoDi Jawa Barat Menggunakan Rapidminer". Dalam penelitian tersebut, melalui penerapan teknik data mining algoritma *K-Means* , peneliti menghitung. Hasil penelitian , yaitu Dengan menggunakan teknik data mining Algoritma K-Means, peneliti menghitung jumlah penduduk miskin di Kota/Kabupaten Jawa Barat berdasarkan atribut yang terdapat, yiatu Kode Kota atau Kabupaten, Nama Kota atau Kabupaten, Jumlah Penduduk Miskin, dan Tahun. Analisis dilakukan terhadap 513 total dataset, menghasilkan *cluster* dengan 196 anggota, *cluster* 1 dengan 38 anggota, cluster 2 dengan 118 anggota dan cluster 3 dengan161 anggota dengan menunjukkan nilai Davies Bouldin-Index yaitu -0,563 sebagai indikasi hasil *cluster* 4 yang dapat diidentifikasi dengan baik.[4]

Tujuan dilakukannya penelitian ini, yaitu:

1. Untuk mengetahui cara pengelompokan tingkat kemiskinan di pulau jawa bedasarkan kabupaten atau kota menggunakan bahasa pemrograman python
2. Untuk mengetahui hasil dari pengelompokan kemiskinan di pulau jawa bedasarkan kabupaten/ kota dengan algoritma k-means clustering
3. Untuk menetukan jumlah cluster yang optimal dengan Davies Bouldien Index
4. Untuk mengetahui penerapan deployment menggunakan framework streamlit
5. Untuk mengetahui hasil penerapan deployment menggunakan framework streamlit

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data Mining suatu tahap pencarian pola atau informasi menarik dari volume data yang besar. Tahapan data mining terdiri dari pengumpulan data, ekstraksi data, analisa data, dan statistik data. Istilah lai juga umum dikenal sebagai knowledge discovery(proses penemuan pengetahuan dari data), knowledge extraction(proses ekstrasi pengetahuan dari data), data/pattern analysis (analisis data dan pola), dan information harvesting(pengambilan informasi dari data. [5].

2.2. *Clustering*

Klasterisasi adalah alat yang kuat yang dapat digunakan untuk menemukan pola dan tren dalam data, Klasterisasi adalah proses penggabungan data perlu dibagi ke dalam sejumlah klaster atau kelompok, sehingga tingkat kemiripan yang tinggi antar data dalam satu kelompok dan rendah antar data kelompok yang berbeda. Klasterisasi dapat digunakan untuk berbagai tujuan, seperti text mining dan analisis web.[6] Menurut penelitian yang dilakukan oleh [7]clustering memiliki beberapa metode yaitu :

1. Hierarchical Clustering
   Metode klasterisasi hirarkis tidak ditentukan jumlah cluster yang diinginkan pada langkah awal sehingga dapat dilihat pembentukan kelompok menggunakan program dendogram.
2. Non-Hierarchical Clusterin
   Metode klasterisasi non hirarkis, total kelompok harus ditentukan terlebih dahulu. Dalam penelitian ini, dua teknik yang diterapkan untuk mengidentifikasi jumlah kelompok melalui elbow plot dan metode silhouette plot.

2.3. *K-Means*

K-Means clustering merupakan suatu teknik pengelompokkan non hierarkis yang digunakan untuk mengaktegorikan data berdasarkan tingkat kemiripan diantara elemen-elemen data tersebut. *K-Means* clustering bekerja dengan cara membagi data kedalam beberapa kelompok, dengan masing-masing kelompok memiliki karakteristik yang sama. Pada K-Means clustering, huruf "k" mewakili jumlah kelompok yang akan dibuat. Nilai "k" ini biasanya ditentukan secara acak.[8]

2.4. Davies Bouldin Index

Davies-Bouldin Index (DBI) adalah suatu metrik evaluasi dalam analisis klaster yang digunakan untuk mengukur kebaikan pembagian klaster pada suatu dataset. DBI mengambil nilai numerik yang menunjukkan seberapa baik klaster terbentuk, dengan mempertimbangkan rata-rata jarak antar pusat klaster dan seberapa tersebar data di dalam klaster. Semakin rendah nilai DBI, semakin baik kualitas klaster, yang menandakan klaster memiliki pemisahan yang baik dan kompak. Sebaliknya, nilai yang tinggi dapat menunjukkan adanya tumpang tindih antar klaster atau ketidakhomogenan dalam pembagian data..[9]

2.5. *Python*

Bahasa pemrograman python adalah bahasa pemrograman yang mengalami perkembangan dan popularitas yang meningkat. Bahasa pemrograman ini memiliki banyak kelebihan yang menjadikannya sebagai alternatif yang sesuai untuk berbagai keperluan. phyton merupakan pemrograman yang dinamis sehingga mudah dibaca. Python dapat digunakan untuk berbagai keperluan, seperti Machine learning dapat digunakan untuk pengembangan algoritma machine learning, seperti klasifikasi, regresi, dan clustering.[10]

3. METODE PERANCANGAN

3.1. Sumber Data

Dalam penelitian tugas akhir ini, seorang peneliti memperoleh data Informasi dan Kemiskinan

pada penelitian ini untuk mengetahui tingkat kemiskinan Kabupaten/Kota di Pulau Jawa pada tahun 2022. Data ini diperoleh dari sumber resmi yaitu Badan Pusat Statistik (BPS), BPS berkantor pusat di Jln. Dr. Sutomo 6-8, Jakarta Pusat 10710. dengan judul "Data dan Informasi Kemiskinan di Indonesia Tahun 2022." Peneliti mengakses data ini melalui situs web resmi *BPS (https://bps.go.id/)* pada tanggal 10 Agustus 2023 dan mengumpulkannya dalam sebuah bentuk tabel excel.

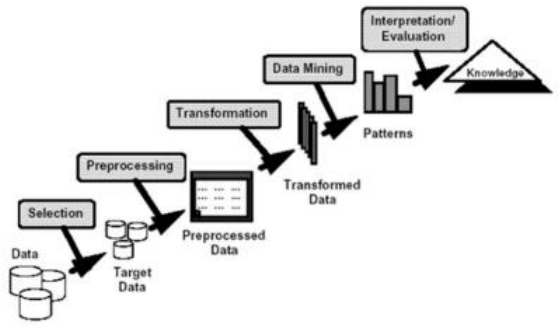
3.2. Teknik Pengumpulan Data

Berikut adalah metode pengumpulan data yang diterapkan oleh peneliti dalam Tugas Akhir ini, yaitu:

1. Pada penelitian ini, metode studi dokumen digunakan untuk mendapatkan dari dokumen dan catatan yang tersedia, dengan sumber data utama adalah situs web resmi Badan Pusat Statistik. Setelah data berhasil diperoleh, langkah berikutnya adalah menganalisis, mengolah, dan memanfaatkan data tersebut oleh peneliti guna mendapatkan informasi yang relevan untuk penelitian ini.
2. Studi literatur merupakan landasan teoritis yang menjadi fondasi penting dalam penelitian Tugas Akhir ini, berperan sebagai panduan utama dalam memecahkan masalah yang dihadapi serta sebagai sumber referensi yang sangat relevan selama proses penelitian. Peneliti merujuk kepada beragam referensi, termasuk jurnal-jurnal nasional dan internasional yang berkaitan erat dengan topik penelitian untuk menguatkan dan mendukung penelitian ini.

Langkah-langkah pengumpulan data yang diterapkan dalam penelitian ini menggunakan data sekunder sesuai dengan sumbernya. Sebanyak 112 data pada tahun 2022 diperoleh dari Bdan Pusat Statistik.

3.3. Tahapan Perancangan



Gambar 1. Tahap KDD

Langkah-langkah dalam merancang penelitian ini melibatkan penerapan *Knowledge Discovery in Database* (KDD), Knowledge Discovery Database (KDD) merupakan proses untuk mengekstrak pengetahuan yang berharga dari kumpulan data dengan volum yang besar. KDD melibatkan berbagai teknik statistik, komputasi, dan kecerdasan buatan untuk menemukan pola, tren, dan hubungan dalam data.meringkas langkah-langkah KDD yang terdiri dari data seleksi, pra-proses data, data transformasi, data mining, interpretasi dan evaluasi.[9]

Dalam proses KDD terdiri dari beberapa tahapan, yaitu:

1. Seleksi Data (*Data Selection*)
   Tahapan awal KDD adalah data selection, yaitu memilih variabel atau atribut yang ada pada data dan informasi tahun 2022 yang akan digunakan dalam analisis K-Means Clustering. Variabel harus relevan dengan masalah yang ingin dipecahkan, yaitu jumlah penduduk miskin, garis kemiskinan, presentase kemiskinan, P1, dan P2.
2. Preprocessing
   Sebelum melakukan proses data mining, diperlukan tahapan pembersihan pada data dan informasi tahun 2022 yang menjadi fokus KDD, Tahap praprocessing bertujuan untuk mempersiapkan data agar siap untuk dianalisis.
3. Data Transformation
   Pada langkah data Data Transformation disesuaikan dan diubah ke dalam format yang sesuai, seperti format .xlsx, untuk memungkinkan penggunaan alat data mining yang memakai algoritma K-Means Clustering. Proses ini bertujuan untuk memastikan bahwa data siap untuk diolah dan dianalisis dengan tepat menggunakan algoritma K-Means Clustering.
4. Data Mining
   Data mining merupakan suatu proses pengumpulan dan pemrosesan data dengan tujuan untuk mengidentifikasi pola, hubungan, atau tren yang berguna dari data. Proses ini dapat dilakukan dengan berbagai teknik, seperti klasifikasi, regresi, clustering, dan outlier detection.
5. Tahap interpretasi dan evaluasi
   Pada tahap ini, hasil data mining akan diubah menjadi informasi yang dapat dipahami. Informasi ini dapat digunakan untuk pengambilan keputusan. Hasil data mining akan dievaluasi menggunakan *Davies-Bouldin Index* (DBI).

4. HASIL DAN PEMBAHASAN

4.1. Penerapan Algortima K-Means untuk Pengelompokan Tingkat Kemiskinan di Pulau Jawa bedasarkan Kabupaten atau kota menggunakan bahasa Pemrograman python

4.1.1. Hasil Pengumpulan Data (Data Collecting)

Data yang diterapkan dalam penelitian ini adalah data Informasi dan Kemiskinan tahun 2022. Data ini mencakup atribut Nama Kabupaten atau Kota, Jumlah Penduduk Miskin (dalam ribuan), Presentase

Penduduk Miskin, P1, P2, dan Garis Kemiskinan. Jumlah total data yang digunakan berjumlah 112 baris.

4.1.2. Data Selection

Pemilihan atribut Data merupakan tahapan dimana data yang tidak diperlukan dihilangkan saat di proses dalam suatu model. Atribut yang ada pada data informasi dan kemiskinan tahun 2022, yaitu Nama Kabupaten atau kota, Jumlah penduduk miskin (000), presentase penduduk miskin, P1, P2, dan Garis Kemiskinan. Seperti ditampilkan pada gambar berikut:

Tabel 1. Data Informasi dan Kemiskinan Tahun 2022 sebelum Selection

V1	V2	V3	V4	V5	V6
Bogor	474,74	7,73	1,49	0,45	443.787
Sukabumi	186,28	7,34	0,93	0,19	357.636
Cianjur	246,81	10,55	1,35	0,27	406.829
…	…	…	….	…	…
Sleman	98,92	7,74	1,18	0,27	450.763
Kota Yogyakarta	29,68	6,62	0,80	0,13	601.905

Keterangan :

a) V1 = kabupaten/kota
b) V2 = jumlah penduduk miskin (000)
c) V3 = presentase penduduk miskin
d) V4 = P1
e) V5 = P2
f) V6 = garis kemiskinan (Rp/Kap/Bulan)

4.1.3. Data *Pre-processing*

Dataset yang diperoleh dari Badan Pusat Statistik Indonesia berupa data informasi dan kemiskinan tahun 2022 terdapat missing value 9 baris pada kolom presentase penduduk miskin dan Garis kemiskinan (Rp/Kap/Bulan) dengan tampilan sebagai berikut:

Tabel 2. Data Informasi dan Kemiskinan tahun 2022 sebelum Pre-Processing

V1	V2	V3	V4	V5	V6
Bogor	474,74	7,73	1,49	0,45	443.787
Sukabumi	186,28	7,34	0,93	0,19	357.636
Cianjur	246,81	10,55	1,35	0,27	406.829
…	…	…	…	…	…
Sleman	98,92	7,74	1,18	0,27	450.763
Kota Yogyakarta	29,68	6,62	0,80	0,13	601.905

Keterangan :

a) V1 = kabupaten/kota
b) V2 = jumlah penduduk miskin (000)
c) V3 = presentase penduduk miskin
d) V4 = P1
e) V5 = P2
f) V6 = garis kemiskinan (Rp/Kap/Bulan)

4.1.4. Data Transformation

Transformasi data Transformasi data merupakan tahapan di mana format data mengalami perubahan. struktur, atau nilai data untuk membuatnya lebih terorganisir dan mudah digunakan oleh manusia maupun komputer. Transformasi data meliputi tugas seperti pembersihan, standarisasi, atau agregasi, yang bertujuan meningkatkan kualitas dan kegunaan data untuk analisis yang lebih akurat.

4.1.5. Data Mining

Data Mining pemodelan K-Means Clustering untuk menentukan tingkat kemiskinan di Pulau Jawa dengan menggunakan bahasa pemrograman pyhton, data mining pada penelitian ini dievaluasi menggunakan metode Davies Bouldin-Index untuk melihat jumlah cluster yang optimal.

Langkah-langkah K-Means Clustering, sebagai berikut:

1. Mengidentifikasi jumlah kelompok, dengan evaluasi menggunakan davies bouldien-index dan diperoleh cluster 2 yang paling optimal atau cluster yang paling baik. Dengan nilai DBI 0,9315.
2. Menentukan Nilai Centroid Awal
   Langkah yang kedua yaitu memilih nilai awal centroid/ titik pusat cluster untuk pemodelan *K-Means.*

Titik awal data yang diperoleh, yaitu:

Tabel 3. Centroid Awal

Centroid	Jumlah Penduduk Miskin(000)	Presentase Penduduk Miskin	P1
C0	74,07	10,79	1,67
C1	276,67	10,42	1,81

3. Perhitungan Jarak Data ke Setiap Klaster dengan Python
4. Setelah menetapkan pusat awal klaster, langkah berikutnya melibatkan perhitungan jarak ke setiap klaster. Dalam menghitung nilai jarak ke masing-masing klaster, peneliti menerapkan rumus Euclidean Distance. Rumus tersebut dirumuskan sebagai berikut:
   Rumus *Euclidean Distance*:

$$D(ij) = \sqrt{(X^1 i - X^1 j)^2 + (X^2 i - X^2 j)^2 + \cdots + (X_k i - X_k j)^2} \quad (10)$$

dimana:
D(i,j) = Jarak data ke *i* ke pusat *cluster j*
Xki = Data ke *i* atribut data ke *k*
Xkj = Titik pusat ke *j* pada atribut ke *k*

5. Iterasi 1
   Pada langkah selanjutnya yaitu Mengalkulasi jarak dari setiap data ke centroid yang paling dekat. Jarak terpendek akan menentukan klaster yang akan diassign oleh data tersebut. Berikut adalah perhitungan jarak ke setiap centroid dengan menggunakan Python. pada data ke1:

Pada Cluster 0 terdapat 89 Data yang tergabung ke dalam nya

Tabel 4. Cluster 0 Iterasi 1

Kabupaten atau kota	JPM(000)	PPM	P1
Ciamis	93,96	7,72	1,07
Kuningan	140,25	12,76	2,14
Majalengka	147,12	11,94	1,55
…	…	…	…
Sleman	98,92	7,74	1,18
Kota Yogyakarta	29,68	6,62	0,8

Pada Cluster 1 terdapat 24 Data yang tergabung ke dalam nya

Tabel 5. Cluster 1 Iterasi 1

Kabupaten atau kota	JPM (000)	PPM	P1
Bogor	474,74	7,73	1,49
Sukabumi	186,28	7,34	0,93
Cianjur	246,81	10,55	1,35
..	…	…	…
Sumenep	206,2	9,27	3,72
Tangerang	270,52	6,92	1,06

Tabel 6. Centroid Baru Iterasi 1

Centroid	JPM(000)	PPM	P1
C0	87, 50191	8, 915169	1,326742
C1	231, 86875	10, 585833	1,908750

Dikarenakan terdapat perpindahan data ke dalam klaster, seperti data dari Kabupaten atau Kota Grobogan dan Kediri, langkah selanjutnya melibatkan perhitungan centroid baru untuk masing-masing klaster berdasarkan data yang termasuk di dalam setiap klaster.

Pada Cluster 0 terdapat 87 Data yang tergabung ke dalam nya

Tabel 7. Cluster 0 Iterasi 2

Kabupaten atau kota	JPM(000)	PPM	P1
Ciamis	93,96	7,72	1,07
Kuningan	140,25	12,76	2,14
Majalengka	147,12	11,94	1,55
..	…	…	…
Sleman	98,92	7,74	1,18
Kota Yogyakarta	29,68	6,62	0,8

Pada Cluster 1 terdapat 26 Data yang tergabung ke dalam nya.

Tabel 8. Cluster 1 Iterasi 2

Kabupaten atau kota	JPM(000)	PPM	P1
Bogor	474,74	7,73	1,49
Sukabumi	186,28	7,34	0,93
Cianjur	246,81	10,55	1,35
…	…	…	…
Sumenep	206,2	9,27	3,72
Tangerang	270,52	6,92	1,06

Tabel 9. Centroid Baru Iterasi 2

Centroid	JPM(000)	PPM	P1
C0	85, 689770	8, 862069	1,315632
C1	226, 827308	10, 635000	1,901154

Sebab pada iterasi ketiga tidak terdapat lagi perpindahan klaster untuk data Kabupaten atau Kota, oleh karena itu, proses iterasi diakhiri dan menghentikan pada iterasi ketiga.Bedasarkan hasil perhitungan jarak *Euclidean Distance* maka diketahui Kabupaten atau kota yang termasuk kedalam dua *cluster* tersebut, berikut tabel mengenai anggota pada setiap clusternya.

4.2. Hasil *K-Means Clustering* Menggunakan bahasa Pemrograman Python

Tabel 10. Anggota Cluster 0 Iterasi

Kabupaten atau kota	JPM(000)	PPM	P1
Ciamis	93,96	7,72	1,07
Kuningan	140,25	12,76	2,14
Majalengka	147,12	11,94	1,55
…	…	…	…
Sleman	98,92	7,74	1,18
Kota Yogyakarta	29,68	6,62	0,8
JUMLAH CLUSTER 0 ADALAH 87 KABUPATEN ATAU KOTA			

Tabel 11. Anggota Cluster 1 Iterasi

Kabupaten atau kota	JPM(000)	PPM	P1
Bogor	474,74	7,73	1,49
Sukabumi	186,28	7,34	0,93
Cianjur	246,81	10,55	1,35
…	…	…	…
Sumenep	206,2	9,27	3,72
Tangerang	270,52	6,92	1,06
JUMLAH CLUSTER 1 ADALAH 26 KABUPATEN ATAU KOTA			

Setelah data diolah menggunakan bahasa pemrograman python, maka dapat dihasilkan pengelompokkan data informasi dan kemiskinan tahun 2022 berdasarkan perhitungan k-means dan implementasinya pada bahasa pemrograman python dapat dilihat pada tabel berikut:

Tabel 12. hasil K-Means

Q1	Q2	Q3	Q4	Q5
Cluster 0	7,88 – 153,40	8,862	1,315	Rendah
Cluster 1	163,2 – 474,74	10,635	1,90	Tinggi

Keterangan :

a) Q1 = Centroid
b) Q2 = JPM (000)
c) Q3 = Rata-rata PPM
d) Q4 = Rata-rata P1
e) Q5 = Kategori Penduduk Miskin

4.3. Evaluasi *K-Means Clustering* dengan Metode Davies Bouldin-Index

Pada penilitian ini evaluasi model *K-Means Clustering* menggunakan metode Davies Bouldin-Index. untuk menentukan cluster yang baik dan optimal agar penelitian ini mendapatkan model yang baik. Nilai Davies Bouldin-Index pada pemodelan K-Means Clustering ini mengahasilkan nilai 0,935 dan *cluster* yang baik diperoleh, yaitu 2 cluster.

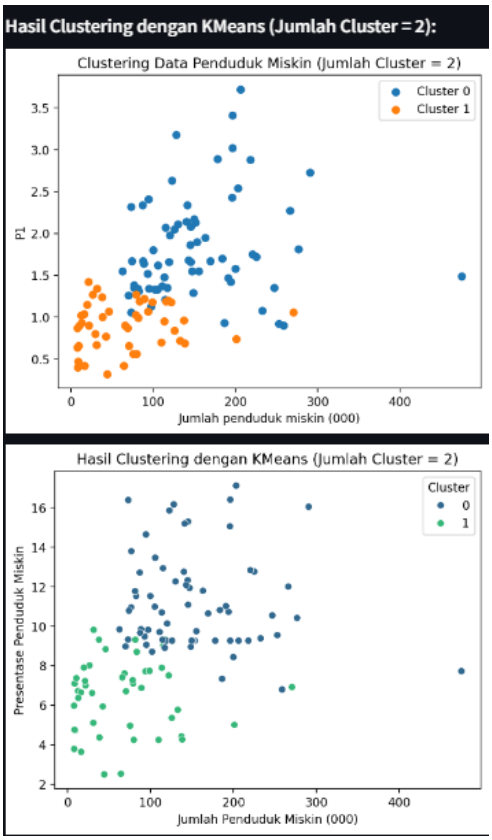
```
In [33]: import numpy as np
         import pandas as pd
         from sklearn.cluster import KMeans
         from sklearn.preprocessing import StandardScaler
         from sklearn.metrics import davies_bouldin_score
         from sklearn.pipeline import make_pipeline

         from sklearn.preprocessing import OneHotEncoder

         # Memuat data
         df = pd.read_excel('Data_Kemiskinan_Pulau_Jawa_cleaned.xlsx')

         # Ekstrak fitur kategorikal
         categorical_features = df.select_dtypes(include=['object']).columns

         # Fitur kategoris encode one-hot
         encoder = OneHotEncoder()
         df_encoded = pd.concat([df.drop(categorical_features, axis=1), pd.DataFrame(encoder.fit_transform(df[categorical_features]).toarr

         # Pilih fitur yang relevan
         selected_features = df_encoded.columns

         # Pilih jumlah cluster (sesuaikan kebutuhan)
         n_clusters = 2

         # Buat alur dengan penskalaan fitur dan KMeans
         pipeline = make_pipeline(StandardScaler(), KMeans(n_clusters=n_clusters, random_state=0))
```

Gambar 2. Metode Davies Bouldin-Index

```
# Ekstrak fitur kategorikal
categorical_features = df.select_dtypes(include=['object']).columns

# Fitur kategoris encode one-hot
encoder = OneHotEncoder()
df_encoded = pd.concat([df.drop(categorical_features, axis=1), pd.DataFrame(encoder.fit_transform(df[categorical_features]).toarr

# Pilih fitur yang relevan
selected_features = df_encoded.columns

# Pilih jumlah cluster (sesuaikan kebutuhan)
n_clusters = 2

# Buat alur dengan penskalaan fitur dan KMeans
pipeline = make_pipeline(StandardScaler(), KMeans(n_clusters=n_clusters, random_state=0))

# Sesuaikan modelnya
pipeline.fit(df_encoded[selected_features])

# Dapatkan label cluster
labels = pipeline.predict(df_encoded[selected_features])

# Hitung Indeks Davies-Bouldin
dbi = davies_bouldin_score(df_encoded[selected_features], labels)
print("Davies-Bouldin Index:", dbi)
```

```
Davies-Bouldin Index: 0.9315749550239337
```

Gambar 3. Metode Davies Bouldin-Index



Gambar 4. Visualisasi K-Means

4.4. Penerapan Deployment Menggunakan Streamlit untuk visualisasi *K-Means Clustering*

Streamlit merupakan library yang disediakan oleh python yang digunakan untuk deployment. Gambar 4 menunjukkan visualisasi hasil pemodelan *K-Means*

Gambar 5. Peta Interaktif Kemiskinan

5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian ini menunjukkan bahwa penggunaan metode clustering dengan bahasa pemrograman Python berhasil mengelompokkan kabupaten atau kota di Pulau Jawa ke dalam dua cluster. Cluster 0 memiliki 87 kabupaten atau kota dengan tingkat kemiskinan terendah, sementara Cluster 1 memiliki 26 kabupaten atau kota dengan tingkat kemiskinan tinggi. Dengan menggunakan metode clustering dengan menggunakan bahasa pemrograma python, penelitian ini berhasil mengelompokkan kabupaten atau kota di Pulau Jawa ke dalam 2 cluster, yaitu pada cluster 0 terdapat 87kabupaten atau kota dengan tingkat kemiskinan terendah dengan nilai rata-rata kemiskinan 8,86%, jumlah penduduk miskin(000) 7,88 – 153,40 dan rata rata nilai P1 1,31%, pada cluster 1 terdapat 26 kabupaten atau kota dengan tingkat kemiskinan tinggi dengan nilai rata-rata kemiskinan sebesar 10,63%, jumlah penduduk miskin(000) 163,2 – 474,74 dan rata-rata nilai P1 1,90%.

Namun, penelitian ini memiliki batasan, seperti fokus pada data tahun 2022 dan parameter tertentu. Untuk menjaga keberlanjutan penelitian, diperlukan pembaruan data berkala dan penelitian lanjutan yang melibatkan faktor-faktor tambahan yang mempengaruhi tingkat kemiskinan.

Saran untuk pemerintah dan penelitian lebih lanjut mencakup perhatian khusus pada wilayah dengan tingkat kemiskinan tinggi, dengan alokasi sumber daya yang lebih besar untuk upaya penanggulangan yang intensif. Program-program khusus, seperti pelatihan keterampilan dan bantuan modal usaha, dapat diimplementasikan untuk meningkatkan taraf hidup masyarakat. Penelitian lebih lanjut dapat melibatkan faktor ekonomi, sosial, dan lingkungan untuk analisis yang lebih komprehensif

dan kebijakan yang lebih efektif. Penggunaan tools lain seperti RapidMiner dapat menjadi pilihan untuk penelitian lanjutan dalam melakukan pengujian.

DAFTAR PUSTAKA

[1] F. Febriansyah and S. Muntari, "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam," 2023.

[2] Ibrahim Sri Hartuti, U. Moonti, and Sudirman, "Pengaruh Tingkat Pendapatan Keluarga Terhadap Kemiskinan Rumah Tangga.," *JOURNAL of ECONOMIC and BUSINESS EDUCATION*, vol. 1, no. 2, May 2023.

[3] S. ' Diyah and E. Adawiyah, "KHIDMAT SOSIAL, Journal of Social Work and Social Service," Apr. 2020.

[4] N. Novitasari, N. D. Nuris, and R. Herdiana, "Jurnal Informatika Terpadu PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA JUMLAH PENDUDUK MISKIN BERDASARKAN KOTA/KABUPATEN DI JAWA BARAT MENGGUNAKAN RAPIDMINER," *Jurnal Informatika Terpadu*, vol. 9, no. 1, pp. 68–73, 2023, [Online]. Available: https://journal.nurulfikri.ac.id/index.php/JIT

[5] I. Gede Iwan Sudipa *et al.*, *DATA MINING*. Padang: PT GLOBAL EKSEKUTIF TEKNOLOGI, 2023. [Online]. Available: www.globaleksekutifteknologi.co.id

[6] Y. Ratna Sari, A. Sudewa, D. Ayu Lestari, and T. Ika Jaya, "PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA KEMISKINAN PROVINSI BANTEN MENGGUNAKAN RAPIDMINER," 2020.

[7] U. Syafiyah, I. Asrafi, B. Wicaksono, D. P. Puspitasari, and M. Sirait, "Analisis Perbandingan Metode Cluster Data Indikator Ketenagakerjaan di Jabar2020," 2020.

[8] A. Z. Irwan and I. Agung Sembe, "Penerapan K-Means Clustering dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM)," 2022. [Online]. Available: http://www.ojs.unm.ac.id/jmathcos

[9] M. Triyana, R. Juita, and C. D. Suhendra, "Penerapan Metode K-Means dalam Pengelompokan Data Penduduk Tidak Mampu di Distrik Oransbari," 2022.

[10] M. Khandava Mulyadien and U. Enri, "Algoritma K-Means Untuk Pengelompokan Bantuan Langsung Tunai (BLT)," *Jurnal Ilmiah Wahana Pendidikan*, vol. 2022, no. 12, pp. 198–210, doi: 10.5281/zenodo.6944517.