

OPTIMALISASI PERFORMA K-MEANS CLUSTERING DENGAN PCA DALAM ANALISIS TINGKAT KEMISKINAN DI JAWA BARAT

Yunita Nurohmah, Rini Mayasari, Betha Nurina Sari

Program Studi Informatika S1, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang
Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang, Indonesia
1910631170056@student.unsika.ac.id

ABSTRAK

Kemiskinan adalah permasalahan kompleks dan sulit diatasi, terutama di Provinsi Jawa Barat yang memiliki tingkat kemiskinan yang tinggi di Indonesia. Data kemiskinan yang akurat menjadi faktor utama dalam mendukung strategi penanggulangan, sementara tingkat kemiskinan suatu wilayah dapat diketahui lebih mudah melalui pengukuran dan penentuan indikator pendukungnya. Penelitian ini bertujuan untuk mengetahui klusterisasi tingkat kemiskinan di Provinsi Jawa Barat dari tahun 2019 hingga 2022 dengan menggunakan metodologi *Knowledge Discovery in Data Base* (KDD). Metode yang diterapkan pada penelitian ini adalah pengelompokan data menggunakan dua algoritma yaitu *Principal Component Analysis* (PCA) dan *K-Means Clustering*, dan menggunakan metode evaluasi *silhouette coefficient* untuk mengetahui nilai dari model. Hasil evaluasi dari 2 skenario model pada penelitian ini menyatakan bahwa membangun model dengan tahap reduksi dimensi variabel terlebih dahulu, lalu dilanjutkan dengan tahap pengelompokan data menggunakan algoritma *K-Means* (skenario 2) yang menghasilkan performa terbaik diantara model lainnya dengan nilai evaluasi 0.74 yang termasuk dalam kriteria kuat.

Kata kunci: *Clustering, K-Means, Kemiskinan, PCA, Silhouette Coefficient*

1. PENDAHULUAN

Indonesia sebagai negara yang sedang berkembang harus memprioritaskan penanggulangan kemiskinan dalam upaya pembangunannya. Meskipun sulit diukur, pengukuran kemiskinan tetap penting sebagai dasar pengambilan kebijakan. Menurut penelitian [1], pengukuran kemiskinan penting untuk menjaga perhatian terhadap orang miskin, mengidentifikasi kebutuhan intervensi, memantau kebijakan yang ditujukan kepada orang miskin, dan mengevaluasi efektivitas pemerintah dalam penanggulangan kemiskinan.

Kemiskinan merupakan masalah kompleks yang sulit diatasi, terutama di Indonesia. Dampak kemiskinan yang berkelanjutan dapat menghambat pembangunan nasional. Kemiskinan adalah ketidakmampuan memenuhi standar hidup minimum dan penuh dengan kekurangan. Pemerintah Indonesia telah menciptakan program-program untuk mengatasi kemiskinan, namun diperlukan solusi yang lebih sistematis dan menyeluruh [2].

Data kemiskinan yang akurat menjadi faktor penting dalam strategi penanggulangan kemiskinan. Pengukuran tingkat kemiskinan suatu wilayah dapat dilakukan dengan menggunakan metode clustering dalam data mining. Metode clustering K-Means adalah teknik analisis data non-hierarki yang membagi data menjadi kelompok berdasarkan karakteristik serupa. Algoritma K-Means efisien dan dapat menghasilkan akurasi yang tinggi dalam pengelompokan data kemiskinan.

Penggunaan teknik *Principal Component Analysis* (PCA) dapat mereduksi dimensi variabel kemiskinan sebelum dilakukan clustering dengan K-

Means. PCA berguna dalam meningkatkan efisiensi model dan menghindari dominasi atribut dengan domain yang lebih besar. Penerapan PCA dalam penelitian ini bertujuan untuk mereduksi dimensi variabel kemiskinan di Jawa Barat dan membandingkan hasil clustering dengan dan tanpa reduksi dimensi [3].

Variabel yang digunakan dalam penelitian ini meliputi garis kemiskinan, indeks keparahan kemiskinan, indeks kedalaman kemiskinan, jumlah penduduk miskin, tingkat pengangguran terbuka, dan upah minimum kabupaten/kota di Jawa Barat. Perbandingan akan dilakukan antara pemodelan clustering dengan K-Means tradisional dan setelah reduksi dimensi menggunakan PCA, dengan mempertimbangkan kualitas cluster yang diukur dengan *Silhouette Coefficient*.

2. TINJAUAN PUSTAKA

2.1. Kemiskinan

Menurut Badan Pusat Statistik (BPS) Indonesia, kemiskinan adalah kondisi di mana seseorang atau rumah tangga tidak memiliki akses yang cukup terhadap kebutuhan dasar seperti makanan, pakaian, dan tempat tinggal yang layak. Kemiskinan dapat diukur berdasarkan tingkat penghasilan, aset, atau konsumsi yang rendah. BPS juga menggunakan indeks kemiskinan yang mencakup kebutuhan dasar manusia seperti makanan, sandang, dan papan. Kemiskinan merupakan masalah kompleks yang melibatkan aspek sosial, ekonomi, budaya, dan lainnya. Rumah tangga miskin kronis sulit keluar dari lingkaran kemiskinan jika menghadapi masalah struktural yang signifikan [4].

2.2. Data Mining

Data mining merupakan proses penambangan informasi dari sekumpulan data dalam *database* yang besar dan kompleks dengan cara mengekstraksi data dan yang tersedia sehingga menjadi pengetahuan dan berguna dalam pengambilan keputusan [5]. *Data mining* digunakan dalam berbagai bidang, seperti bisnis, ilmu pengetahuan, teknologi informasi, dan sebagainya untuk menemukan pola dalam data [6].

2.3. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah kumpulan tahapan dalam proses identifikasi data menjadi kumpulan informasi yang mengandung pola. Informasi ini terdapat dalam basis data berukuran besar yang sebelumnya belum diketahui manfaat potensial dari data tersebut [7]. Tahapan dari KDD diantaranya yaitu *data selection*, *data preprocessing*, *transformation*, *data mining*, dan *evaluation* [8].

2.4. Clustering

Clustering adalah pengelompokan objek berupa data dengan karakteristik atau sifat yang mirip ke dalam sebuah *cluster* dan objek atau data dengan tingkat kemiripan yang rendah terhadap *cluster* yang lain dengan melakukan pemaksimalan kesamaan antar objek dalam sebuah klaster dan meminimalkan kesamaan antar *cluster* [9]. *Clustering* merupakan metode atau teknik menambang data (*data mining*) yang digunakan untuk menganalisa data [10].

2.5. K-Means

K-Means merupakan sebuah metode dalam penambangan data yang bermanfaat untuk mengelompokkan data yang tidak memiliki tingkatan atau hanya bersifat partisi. Metode ini dapat membagi objek menjadi kelompok-kelompok, yang dapat membagi data menjadi satu kelompok yang mempunyai kemiripan dan memisahkannya dengan kelompok lain. Metode ini didasarkan terhadap jarak antar data di dalam *cluster* yang terbentuk. Semakin kecil jarak maka semakin mirip antara data satu dengan data lainnya. Algoritma K-Means memiliki beberapa kelebihan yaitu mampu mengagregasi data dalam objek yang besar dengan proses yang singkat dan efisien [11].

2.6. Metode Elbow

Metode *elbow* adalah metode untuk mengoptimalkan pemilihan titik *cluster* atau *centroid* dalam proses penambangan data. Optimalisasi dilakukan dengan memilih nilai *k* yang kecil dan yang memiliki nilai *witness* rendah. Pada metode *elbow cluster* optimal akan mengalami penurunan nilai SSE yang besar setelah itu titik-titik berikutnya penurunan akan cenderung stabil [12].

2.7. Silhouette Coefficient

Jarak antar data pada *silhouette coefficient* dihitung menggunakan rumus *euclidian distance* [13].

Kriteria pengukuran pengelompokkan berdasarkan score *silhouette coefficient* disajikan pada Tabel 1.

Tabel 1. Kriteria Pengukuran *Silhouette Coefficient*

Score	Kriteria
0.71-1.00	Struktur Kuat
0.51-0.70	Struktur Baik
0.26-0.50	Struktur Lemah
≤ 0.25	Struktur Buruk

2.8. Python

Python merupakan bahasa pemrograman interpreter, interaktif, dan object-oriented dengan sintaks yang jelas dan elegan. Cocok digunakan di berbagai sistem operasi dan didukung modul-modul dengan struktur data tingkat tinggi. Source code Python dikompilasi menjadi byte code dan bisa dieksekusi. Python gratis, bebas digunakan, dan dilengkapi dengan fasilitas seperti debugger, profiler, GUI, dan basis data [14].

2.9. Quantum Geographic Information System

Sistem Informasi Geografis (SIG) adalah sebuah sistem yang berguna untuk proses pengumpulan, pemeriksaan, dan analisis informasi yang berkaitan dengan kondisi geografis permukaan bumi. *Quantum Geographic Information System* (QGIS) adalah *software* SIG secara terbuka dan tidak memerlukan biaya yang biasanya ditujukan untuk pengolahan data geospasial [15].

3. METODE PENELITIAN

Metodologi penelitian ini yaitu menggunakan metode *Knowledge Discovery in Database* (KDD). KDD merupakan kumpulan tahapan dalam proses identifikasi data menjadi kumpulan informasi yang mengandung pola. Informasi ini terdapat dalam basis data berukuran besar yang sebelumnya belum diketahui manfaat potensial dari data tersebut. Proses yang dilakukan yaitu melalui:

3.1. Data Selection

Pada tahap ini dilakukan pemilihan data yang akan diolah. Penelitian ini menggunakan data kemiskinan Kabupaten/Kota di Jawa Barat dari tahun 2019 hingga 2022. Data tersebut diperoleh dari dua sumber utama, yaitu *website* BPS (Badan Pusat Statistik) dan Open Data Jawa Barat. Total terdapat 108 *entry* data mentah yang akan diolah dalam penelitian ini untuk menghasilkan informasi baru yang relevan. Data yang digunakan terdiri dari beberapa variabel, antara lain No, Nama Wilayah, GK (Garis Kemiskinan), IDK (Indeks Kedalaman Kemiskinan), IPK (Indeks Keparahan Kemiskinan), JPM (Jumlah Penduduk Miskin), TPT (Tingkat Pengangguran Terbuka), UMK (Upah Minimum Kabupaten/Kota), dan Tahun.

3.2. Preprocessing and Data Cleaning

Pada tahapan sebelumnya, dibangun sebuah *data set* yang kemudian dilakukan pembersihan (*cleaning*) untuk meningkatkan kualitasnya. Dalam proses pembersihan tersebut, kolom-kolom yang tidak dibutuhkan dalam proses selanjutnya dihapus. Selain itu, dilakukan pengecekan terhadap *missing value* dan data duplikat guna memastikan integritas data yang lebih baik.

3.3. Transformation

Pada tahap ini, dilakukan normalisasi dengan menggunakan *Min-max normalization* yang dimana dilakukan transformasi linier pada data asli dengan melakukan standarisasi data agar nilainya berada dalam rentang 0 hingga 1, di mana nilai terkecil direpresentasikan sebagai 0 dan nilai terbesar direpresentasikan sebagai 1. Dengan menggunakan metode ini, data-data tersebut dapat memiliki keseimbangan yang sebanding antara satu dengan yang lainnya.

3.4. Data Mining

Data mining diimplementasikan dengan dua skenario, yang pertama pemodelan hanya menggunakan *K-Means Clustering* dan yang kedua menggunakan *PCA + K-means*. Kedua skenario tersebut dibangun menggunakan *tools* Google Colaboratory dan menggunakan bahasa Python.

3.5. Interpretation And Evaluation

Setelah didapatkan pemodelan di tahap *data mining* maka selanjutnya akan dilakukan uji coba performa model dengan melihat nilai evaluasi *Silhouette Coefficient* dari kedua skenario model.

3.6. Discovery Knowledge

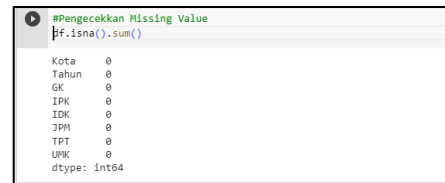
Setelah tahap modeling, hasil dari cluster tiap kabupaten kota di Jawa Barat dapat divisualisasi dalam bentuk pemetaan menggunakan bantuan tools sistem informasi grafis yaitu QGIS. Tujuan visualisasi dalam bentuk pemetaan ini yaitu supaya hasil cluster dapat lebih mudah dilihat dan dipahami. Dimana, nantinya setiap wilayah kabupaten dan kota direpresentasikan dengan warna yang berbeda sesuai dengan karakteristik cluster yang didapat pada tahap modeling.

4. HASIL DAN PEMBAHASAN

Pada tahap *data selection* dilakukan proses pengumpulan data mentah atau *dataset* awal. Pada tahap ini dilakukan pengumpulan data, *dataset* yang digunakan merupakan data publik yang diambil dari *website* opendata.jabarprov.go.id dan *website* BPS.

Pada tahap selanjutnya yaitu data *preprocessing* atau pemrosesan. Langkah pertama adalah dengan melakukan penghapusan pada data yang tidak diperlukan. Data dianggap tidak diperlukan karena tidak berhubungan dengan penelitian atau atribut tersebut tidak berpengaruh terhadap proses analisis

data. Atribut yang dihapus yaitu atribut *id*, *kode_provinsi*, *nama_provinsi*, *kode_kabupaten_kota*, *satuan*, dan atribut *tahun*. Setelah data diintegrasikan, Langkah selanjutnya yaitu melakukan pengecekan *missing value*.

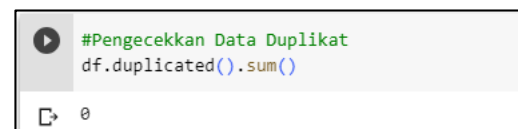


```
#Pengecekan Missing Value
df.isna().sum()

Kota      0
Tahun     0
GK        0
IPK       0
IDK       0
JPM       0
TPT       0
UMR       0
dtype: int64
```

Gambar 1. Pengecekan Missing Value

Pada Gambar 1. menunjukkan bahwa tidak terdapat data yang memiliki nilai kosong. Maka dari itu dapat dilanjutkan ke pengecekan selanjutnya yaitu pengecekan redundansi data atau data yang terindikasi memiliki duplikat.



```
#Pengecekan Data Duplikat
df.duplicated().sum()

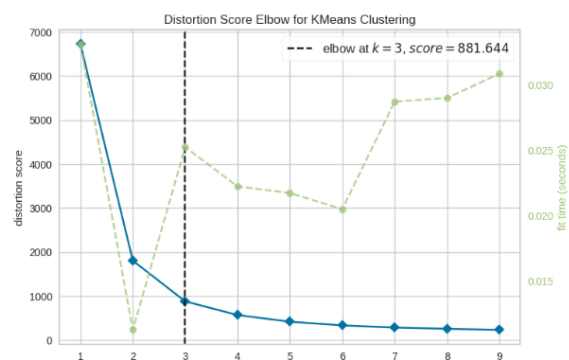
0
```

Gambar 2. Pengecekan Data Duplikat

Setelah data siap diolah, tahapan selanjutnya yaitu proses data mining untuk melihat pola yang terdapat pada data. Teknik yang diterapkan pada proses ini adalah teknik *clustering* yaitu melakukan pengelompokan data berdasarkan tingkat kemiripan dalam *cluster*. terdapat dua skenario yaitu skenario pertama menggunakan algoritma yang digunakan yaitu algoritma *K-Means* tradisional, dan skenario yang kedua menggunakan algoritma *K-Means* dengan optimalisasi menggunakan *PCA* dengan melihat nilai evaluasi *Silhouette* yang terbaik.

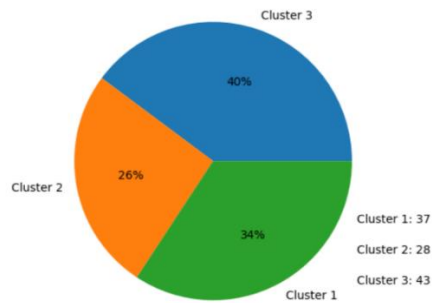
4.1. Skenario Satu (K-Means Tradisional)

Skenario satu menghasilkan tiga *cluster* berdasarkan Kabupaten/kota di Provinsi Jawa Barat dengan nilai *K* optimal yang dihasilkan dari metode elbow yang direpresentasikan pada gambar 3



Gambar 3. Hasil *K* optimal dengan Elbow

Kemudian dilakukan modeling dengan algoritma *K-Means* dan menunjukkan bahwa dari data kemiskinan di Jawa Barat pada tahun 2019-2022 yang dibuat menjadi tiga *cluster* yaitu cluster 1 dengan jumlah 37 anggota, kemudian cluster 2 dengan 28 anggota dan cluster 3 dengan jumlah 43 anggota. Jumlah anggota hasil cluster dapat dilihat pada Gambar 4.



Gambar 4. Visualisas Hasil *Clustering K-Means*

Karakteristik *cluster* pada Tabel 2 digunakan untuk menjelaskan karakter dari setiap cluster yang dihasilkan oleh algoritma *K-Means clustering* dengan melihat dari nilai rata-rata setiap variabel dalam kasus kemiskinan di Jawa Barat.

Tabel 2. Karakteristik *Cluster* Skenario Satu

Cluster	Deskripsi
1	<i>Cluster</i> ini memiliki karakteristik yaitu rata-rata Garis Kemiskinan (GK) yang relatif rendah, namun IPK dan IDK berada pada tingkat yang lebih tinggi dibandingkan dengan <i>cluster</i> lainnya. UMK (Upah Minimum Kabupaten/Kota) di <i>cluster</i> ini berada pada kisaran menengah.
2	<i>Cluster</i> ini memiliki karakteristik yaitu rata-rata Garis Kemiskinan (GK) yang lebih tinggi dibandingkan dengan <i>cluster</i> lainnya. Namun, Indeks Keparahan Kemiskinan (IPK) dan Indeks Kedalaman Kemiskinan (IDK) berada pada tingkat yang rendah. Jumlah Jumlah Penduduk Miskin (JPM) rata-rata dan total pengeluaran transportasi (TPT) berada pada tingkat yang relatif tinggi. UMK (Upah Minimum Kabupaten/Kota) di <i>cluster</i> ini adalah yang tertinggi di antara ketiga <i>cluster</i> .
3	<i>Cluster</i> ini memiliki karakteristik yaitu rata-rata Garis Kemiskinan (GK) yang lebih rendah dibandingkan dengan <i>cluster</i> 2, namun IPK dan IDK berada pada tingkat yang lebih tinggi. (UMK (Upah Minimum Kabupaten/Kota) di <i>cluster</i> ini berada pada kisaran menengah.

Tabel 3. menunjukkan hasil *clustering* Seluruh Kabupaten/Kota di Jawa Barat pada tahun 2019 yang merupakan anggota dari *cluster* 3 yang berarti memiliki karakteristik yang sama.

Tabel 3. Hasil *Cluster* Skenario Satu Tahun 2019

Cluster	Nama Kabupaten/Kota
3	Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Bandung, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.

Tabel 4. menunjukkan hasil *clustering* Kabupaten/Kota di Jawa Barat pada tahun 2020 terbagi menjadi 3 *cluster*. *Cluster* 1 yang ditempati oleh 5 Kabupaten/Kota. Kemudian *cluster* 2 yang ditempati oleh 7 Kabupaten/Kota. Sedangkan untuk *cluster* 3 ditempati oleh 15 Kabupaten/Kota.

Tabel 4. Hasil *Cluster* Skenario Satu Tahun 2020

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

Tabel 5. menunjukkan hasil *clustering* Kabupaten/Kota di Jawa Barat pada tahun 2021 terbagi menjadi 3 *cluster*. *Cluster* 1 yang ditempati oleh 15 Kabupaten/Kota. Kemudian *cluster* 2 yang ditempati oleh 11 Kabupaten/Kota. Sedangkan untuk *cluster* 3 ditempati oleh satu Kabupaten/Kota.

Tabel 5. Hasil *Cluster* Skenario Satu Tahun 2021

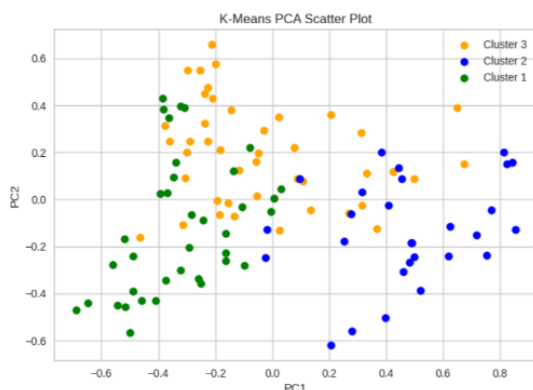
Cluster	Nama Kabupaten/Kota
1	Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Pangandaran, Kota Sukabumi, Kota Cirebon, Kota Tasikmalaya, Kota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

Tabel 6. menunjukkan hasil clustering Kabupaten/Kota di Jawa Barat pada tahun 2022 terbagi menjadi 2 cluster. Cluster 1 yang ditempati oleh 17 Kabupaten/Kota. Sedangkan cluster 2 yang ditempati oleh 10 Kabupaten/Kota.

Tabel 6. Hasil *Cluster* Skenario Satu Tahun 2022

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

Berdasarkan Gambar 5 dapat dilihat Cluster 1 terletak di sebagian besar bagian kiri bawah plot, Cluster 2 terletak di sebagian besar bagian kanan bawah plot dan Cluster 3 terletak di sebagian besar bagian atas plot. Terdapat beberapa titik yang saling berdekatan dalam setiap cluster menunjukkan adanya kemiripan dalam karakteristik di setiap clusternya. Terdapat beberapa titik yang tersebar secara acak di seluruh plot, menunjukkan adanya variasi dalam karakteristik di setiap clusternya. Terdapat overlap antara cluster 1 dan cluster 3 di sebagian wilayah tengah plot, menunjukkan adanya kemungkinan adanya kesamaan atau keragaman dalam fitur-fitur yang dimiliki oleh kedua cluster tersebut.

Gambar 5. *Scatter Plot* Hasil *Clustering* Skenario Satu

4.2. Skenario Dua (PCA + K-Means)

Pada skenario kedua, langkah pertama untuk pemodelan yaitu menggunakan algoritma PCA terlebih dahulu untuk mereduksi variabel menjadi dua dimensi.

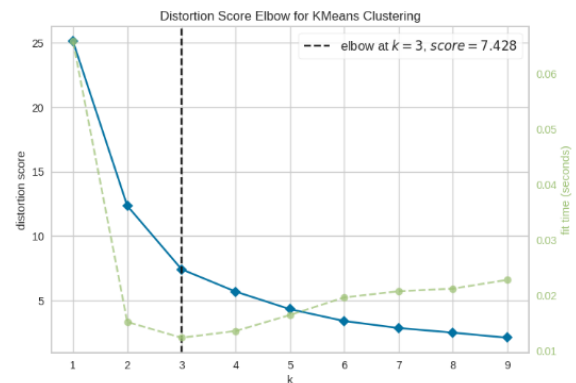
Nilai pada Tabel 7 merupakan hasil dari proses PCA yang telah dilakukan pada Gambar 4.16. Nilai

tersebut mewakili enam variabel pada dataset kemiskinan yang telah melalui proses reduksi menjadi dua dimensi

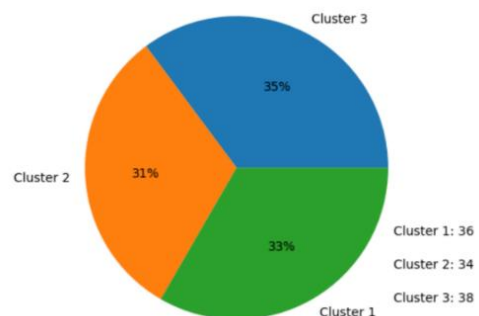
Tabel 7. Hasil Reduksi Dimensi Variabel Menggunakan PCA

PC1	PC2
0.2698260909679	-0.0582823338748514
0.0233614277573971	0.3500630163670091
-0.02883295209887942	0.292193347023774
...	...
-0.48927492469516015	-0.3908821492584138
-0.32160431577293275	0.3960499576038546

Tahap selanjutnya mencari K optimal menggunakan metode *elbow* untuk melakukan *clustering* pada data yang telah melalui proses PCA menggunakan algoritma *K-Means*. Berdasarkan pencarian K optimal diatas, jumlah *cluster* yang digunakan adalah $K=3$ untuk pengelompokan data kemiskinan wilayah di Jawa Barat.

Gambar 6. Hasil K Optimal dengan *Elbow* Skenario Dua

Kemudian dilakukan modeling dengan algoritma *K-Means* dan menunjukkan bahwa dari data kemiskinan di Jawa Barat pada tahun 2019-2022 yang dibuat menjadi tiga *cluster* yaitu cluster 1 dengan jumlah 36 anggota, kemudian cluster 2 dengan 34 anggota dan cluster 3 dengan jumlah 38 anggota. Jumlah anggota hasil cluster dapat dilihat pada Gambar 7.

Gambar 7. Visualisas Hasil *Clustering* PCA + *K-Means*

Karakteristik *cluster* pada Tabel 8 digunakan untuk menjelaskan karakter dari setiap *cluster* yang dihasilkan oleh algoritma *K-Means clustering* dengan melihat dari nilai rata-rata setiap variabel dalam kasus kemiskinan di Jawa Barat.

Tabel 8. Karakteristik *Cluster* Skenario Satu

Cluster	Deskripsi
1	<i>Cluster</i> ini memiliki karakteristik yaitu rata-rata Jumlah Penduduk Miskin yang relatif rendah, namun, IPK dan IDK berada pada tingkat yang lebih tinggi jika dibandingkan dengan <i>cluster</i> lainnya. Garis Kemiskinan (GK) dan Tingkat Pengangguran Terbuka (TPT) berada pada tingkat yang stabil. Serta Upah Minimum Kabupaten/Kota (UMK) di <i>cluster</i> ini berada pada kisaran menengah.
2	<i>Cluster</i> ini memiliki karakteristik yaitu rata-rata Garis Kemiskinan (GK), Jumlah Penduduk Miskin (JPM), Tingkat Pengangguran Terbuka (TPT) dan Upah Minimum Kabupaten/Kota (UMK) yang lebih tinggi dibandingkan dengan <i>cluster</i> yang lain. Namun, IPK dan IDK yang relatif stabil.
3	<i>Cluster</i> ini memiliki karakteristik yaitu rata-rata Garis Kemiskinan (GK), Indeks Keparahisan Kemiskinan (IPK), Tingkat Pengangguran Terbuka (TPT), dan Upah Minimum Kabupaten/Kota (UMK) yang lebih rendah dibandingkan dengan <i>Cluster</i> 1 dan 2. Indeks Kedalaman Kemiskinan (IDK) serta Jumlah Penduduk Miskin (JPM) di <i>cluster</i> 3 berada pada kisaran menengah.

Tabel 9. menunjukkan hasil *clustering* Kabupaten/Kota di Jawa Barat pada tahun 2019 terbagi menjadi 3 *cluster*. *Cluster* 1 yang ditempati oleh 4 Kabupaten/Kota. Kemudian *cluster* 2 yang ditempati oleh 8 Kabupaten/Kota. Sedangkan untuk *cluster* 3 ditempati oleh 15 Kabupaten/Kota.

Tabel 9. Hasil *Cluster* Skenario Dua 2019

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

Tabel 10 menunjukkan hasil *clustering* Kabupaten/Kota di Jawa Barat pada tahun 2020 terbagi menjadi 3 *cluster*. *Cluster* 1 yang ditempati oleh 8 Kabupaten/Kota. Kemudian *cluster* 2 yang

ditempati oleh 9 Kabupaten/Kota. Sedangkan untuk *cluster* 3 ditempati oleh 10 Kabupaten/Kota.

Tabel 10. Hasil *Cluster* Skenario Dua 2020

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

Tabel 11 menunjukkan hasil *clustering* Kabupaten/Kota di Jawa Barat pada tahun 2021 terbagi menjadi 3 *cluster*. *Cluster* 1 yang ditempati oleh 14 Kabupaten/Kota. Kemudian *cluster* 2 yang ditempati oleh 9 Kabupaten/Kota. Sedangkan untuk *cluster* 3 ditempati oleh 4 Kabupaten/Kota.

Tabel 11. Hasil *Cluster* Skenario Dua 2021

Cluster	Nama Kabupaten/Kota
1	Kabupaten Bogor, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Bandung Barat, Kota Tasikmalaya.
2	Kabupaten Bogor, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kota Bogor, Kota Bandung, Kota Bekasi, Kota Depok, Kota Cimahi.
3	Kabupaten Sukabumi, Kabupaten Ciamis, Kabupaten Pangandaran, Kota Banjar.

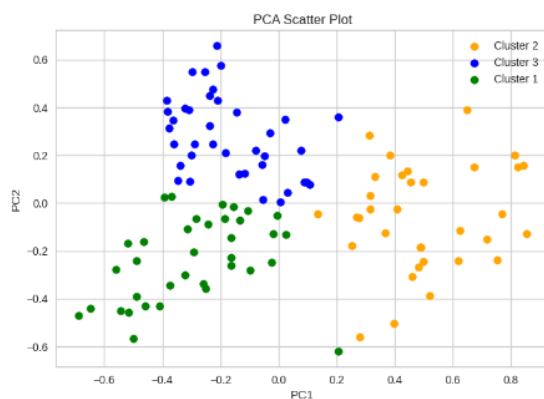
Tabel 12 menunjukkan hasil *clustering* Kabupaten/Kota di Jawa Barat pada tahun 2020 terbagi menjadi 3 *cluster*. *Cluster* 1 yang ditempati oleh 8 Kabupaten/Kota. Kemudian *cluster* 2 yang ditempati oleh 9 Kabupaten/Kota. Sedangkan untuk *cluster* 3 ditempati oleh 10 Kabupaten/Kota.

Tabel 12. Hasil *Cluster* Skenario Dua 2022

Cluster	Nama Kabupaten/Kota
1	Kabupaten Bogor Kabupaten Cianjur Kabupaten Garut Kabupaten Kuningan Kabupaten Cirebon Kabupaten Sumedang Kabupaten Indramayu Kabupaten Subang Kabupaten Bandung Barat Kota Tasikmalaya
2	Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kota Bogor, Kota Bandung, Kota Bekasi, Kota Depok, Kota Cimahi.
3	Kabupaten Sukabumi, Kabupaten Bandung, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Majalengka, Kabupaten

	Pangandaran, Kota Sukabumi, Kota Cirebon, Kota Banjar.
--	---

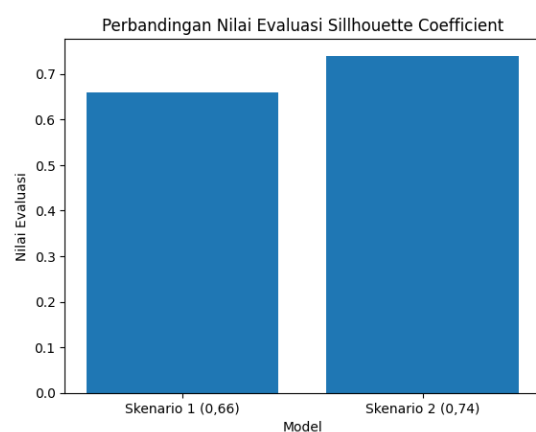
Berdasarkan Gambar 7, *Cluster 1* terletak dibagian kiri bawah plot, *cluster 2* sebagian terletak ditengah atas plot dan *Cluster 3* terletak di kiri atas sebagian di tengah plot. Pada plot ini hanya beberapa titik yang bersinggungan antara *Cluster 1*, *Cluster 2* dan *Cluster 3*. Pada plot diatas juga sudah tidak terdapat titik-titik yang tersebar secara acak dimana berarti proses PCA berhasil membuat karakteristik masing-masing *Cluster* semakin jelas. Selain itu, sudah tidak terdapat overlap antara cluster 1 dan cluster 2 di sebagian wilayah tengah plot yang sebelumnya terlihat dengan jelas.



Gambar 7. Scatter Plot Hasil Clustering PCA + K-Means

4.3. Evaluasi Model

Berdasarkan hasil penilaian terhadap skenario satu dan dua, model yang dibuat menggunakan skenario dua menghasilkan nilai evaluasi yang lebih optimal, hal ini dibuktikan oleh hasil evaluasi pada Gambar 8.

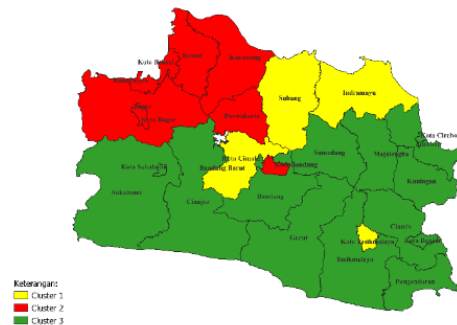


Gambar 8. Perbandingan Nilai Evaluasi Dua Skenario

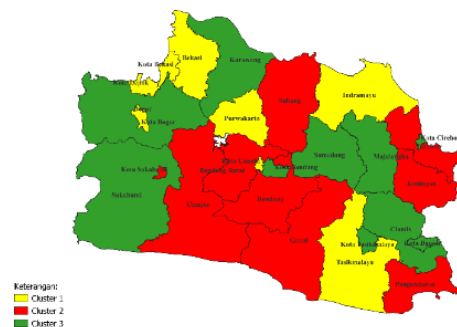
Dari hasil tersebut dapat disimpulkan bahwa pada tahap visualisasi akan menggunakan skenario yang kedua.

4.4. Visualisasi

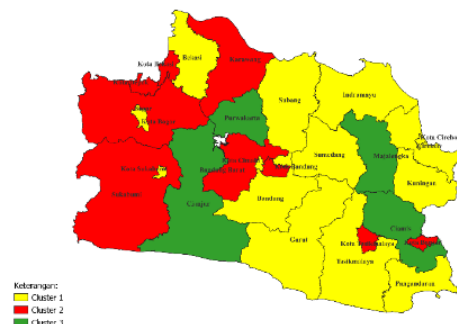
Peta Daerah Kemiskinan Provinsi Jawa Barat 2019



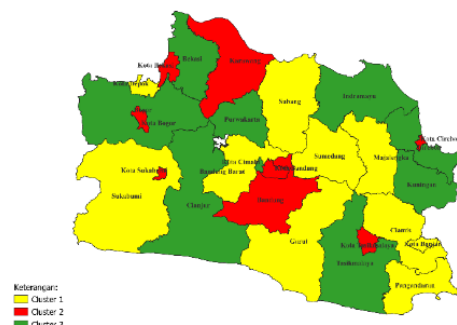
Peta Daerah Kemiskinan Provinsi Jawa Barat 2020



Peta Daerah Kemiskinan Provinsi Jawa Barat 2021



Peta Daerah Kemiskinan Provinsi Jawa Barat 2022



Gambar 9. Peta Daerah Miskin Jawa Barat 2019-2022

Dalam rentang waktu tahun 2019 hingga 2022, terdapat perbedaan yang signifikan dalam perubahan tingkat kemiskinan di setiap Kota/Kabupaten. Beberapa Kota/Kabupaten mengalami perubahan *cluster* selama periode tersebut. Contohnya, Kabupaten Bandung Barat, Kabupaten Cimahi, Kabupaten Bandung, Kabupaten Cianjur, Kabupaten Kuningan, dan Kabupaten Cirebon pada tahun 2019 berada dalam *Cluster* 3, namun mengalami kenaikan tingkat kemiskinan hingga berpindah ke *Cluster* 1 pada tahun 2020. Selanjutnya, rata-rata tingkat kemiskinan mulai membaik pada tahun 2021-2022. Kota Bandung juga mengalami perubahan serupa, dari *Cluster* 3 di tahun 2019 menjadi *Cluster* 1 pada tahun 2021, namun kembali ke *Cluster* 3 di tahun 2022.

Di beberapa Kota/Kabupaten, terjadi keadaan yang berbeda pada tahun 2020, di mana tingkat kemiskinan yang sebelumnya menurun di tahun 2019, kembali naik dan berada dalam *Cluster* 3. Contohnya, Kota Bogor, Kabupaten Bogor, Kota Depok, Kota Bekasi, Kabupaten Bekasi, Kabupaten Karawang, dan Kabupaten Purwakarta. Namun, pada tahun 2021, keadaan mulai membaik, termasuk Kota Sukabumi dan Kabupaten Sukabumi yang kembali ke *Cluster* 2. Namun, di tahun 2022, hanya Kota Bogor, Kota Bekasi, dan Kota Karawang yang tetap berada di *Cluster* 2, sementara Kota/Kabupaten lainnya mayoritas berada di *Cluster* 1 dan *Cluster* 3.

Selain itu, Kabupaten Tasikmalaya, Kabupaten Pangandaran, Kabupaten Ciamis, Kota Bandar, Kabupaten Indramayu, dan Kabupaten Majalengka juga mengalami perubahan yang signifikan. Pada tahun 2020, kebanyakan daerah tersebut mengalami peningkatan tingkat kemiskinan, namun mulai membaik sejak tahun 2021 hingga 2022. Namun, Kota Tasikmalaya memiliki pola yang berbeda, dengan perubahan dari *Cluster* 1 pada tahun 2019, menjadi *Cluster* 3 pada tahun 2020, dan kembali memburuk di tahun 2021 dan 2022 dengan berada di *Cluster* 3.

5. KESIMPULAN DAN SARAN

Penerapan teknik *data mining* dengan optimalisasi performa *K-Means clustering* menggunakan algoritma PCA untuk pengelompokan daerah miskin berdasarkan Kabupaten/Kota di Jawa Barat pada tahun 2019-2022 menghasilkan tiga cluster dalam masing-masing skenario. Jumlah *cluster* yang digunakan adalah tiga berdasarkan metode *elbow*. Kedua, evaluasi pengelompokan daerah miskin menggunakan metode *silhouette coefficient* menunjukkan nilai 0,61 dalam skenario satu (layak) dan 0,74 dalam skenario dua (kuat). Ketiga, visualisasi pemetaan daerah kemiskinan di Provinsi Jawa Barat dari tahun 2019-2022 menggunakan QGIS menunjukkan perubahan signifikan dalam tingkat kemiskinan. Beberapa daerah, seperti Kabupaten Bandung Barat, Cimahi, Bandung, Cianjur, Kuningan, dan Cirebon, berpindah dari *Cluster* 3 ke *Cluster* 1 pada tahun 2020 dan mengalami perbaikan pada tahun 2021-2022. Kota Bandung juga mengalami perubahan

serupa. Terdapat juga perubahan bertolak belakang di beberapa daerah pada tahun 2020, namun mayoritas daerah berada dalam *Cluster* 1 dan 3.

DAFTAR PUSTAKA

- [1] K. Berliani, "Pengaruh Tingkat Pengangguran, Tingkat Pendidikan dan Laju Pertumbuhan Penduduk Terhadap Tingkat Kemiskinan Penduduk Provinsi Jawa Barat Tahun 2015-2020," *Syntax Lit. ; J. Ilm. Indones.*, vol. 6, no. 2, p. 872, 2021, doi: 10.36418/syntax-literate.v6i2.2244.
- [2] S. Kuncoro, A., & Harmadi, *Mozaik Demografi Untaian Pemikiran tentang Kependudukan dan Pembangunan*. Jakarta: Salemba Emot, 2016.
- [3] B. Dash, D. Mishra, A. Rath, and M. Acharya, "A hybridized K-means clustering approach for high dimensional dataset," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 2, pp. 59–66, 2010, doi: 10.4314/ijest.v2i2.59139.
- [4] H. Jacobus, Elvira, P. Kindangen, and N. Walewangko, Een, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI KEMISKINAN RUMAH TANGGA DI SULAWESI UTARA," vol. 19, no. 3, pp. 86–103, 2018.
- [5] J. Suntoro, *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Elex Media Komputindo, 2019.
- [6] M. R. Nahjan, N. Heryana, and A. Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," *J. Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 1–4, 2023.
- [7] Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, "Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.
- [8] A. Fahllevi, "Proses Data Mining KDD," *Binus University*, 2022. Pada Pengguna Brand Hp Menggunakan Metode K-Means," *Pros. Semin. Nas. Ilmu Komput.*, vol. 1, no. 1, p. 2021, 2021.
- [10] B. Nurseptia, A. Voutama, N. Haeryana, and J. HSRonggo Waluyo, "Penerapan Algoritma K-Means Untuk Mengelompokkan Kabupaten/Kota Dalam Upaya Pemetaan Lapangan Pekerjaan Baru," *J. Teknol. Informasi*, vol. 6, no. 2, pp. 181–186, 2022.
- [11] A. P. Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 164, 2023, doi: 10.53513/jsk.v6i1.7351.
- [12] A. Ardiansyah and F. Nggeboe, "Kajian Yuridis Penelantaran Anak Oleh Orang Tua Menurut Perspektif Hukum Indonesia," *Leg. J. Huk.*, vol.

- 10, no. 1, p. 146, 2019, doi: 10.33087/legalitas.v10i1.160.
- [13] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.
- [14] T. Wahyono, *Fundamental Of Python For Machine Learning*. Penerbit Gava Media, 2021.
- [15] S. Bahri, D. M. Midyanti, and R. Hidayati, "Pemanfaatan QGIS Untuk Pemetaan Fasilitas Layanan Masyarakat Di Kota Pontianak," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 1, p. 70, 2020, doi: 10.24114/cess.v5i1.15666.