

Terbit online pada laman web jurnal: <https://jurnal.plb.ac.id/index.php/tematik/index>

T E M A T I K

Jurnal Teknologi Informasi Komunikasi (e-Journal)

Vol. 9 No. 1 (2022) 29 - 35

ISSN Media Elektronik: 2443-3640

Pengelompokan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient *West Java Province Poverty Data Grouping Using the K-Means Algorithm with Silhouette Coefficient*

Nabila Nur Fransiska R¹, Dwi Suci Angraeni², Ultach Enri³^{1,2,3}Teknik Informatika, Ilmu Komputer, Universitas Singaperbangsa Karawang¹nabila.fransiska18150@student.unsika.ac.id, ²dwi.suci18191@student.unsika.ac.id*, ³ultach@staff.unsika.ac.id***Abstract**

Serious poverty is still one of the problems in Indonesia, especially in West Java Province. The level of underdevelopment and unemployment is still the basis for poverty. Poverty in each region is certainly different. The government needs to know which areas fall into the categories of high poverty levels and low poverty levels so that they can make solutions to set priorities in providing assistance. Therefore, a data mining technique is needed that can classify the poverty level of areas in West Java, namely the clustering technique with the K-Means algorithm. The purpose of this research is to classify poverty data in West Java Province so that it can be used as information to determine the right policy to distribute aid to the community from the West Java government. The results obtained based on the test, the clusters obtained were 2 clusters with cluster 0 of high poverty level as many as 14 regions and cluster 1 of low poverty level as many as 13 regions. Based on the test, the K-Means Algorithm obtains a Silhouette Coefficient of 0.576 and is included in the medium structure category. With the results of grouping poverty data, the government can channel aid more precisely.

Keywords: data mining, k-means, poverty

Abstrak

Kemiskinan yang cukup serius masih menjadi salah satu masalah yang ada di Indonesia, terutama di Provinsi Jawa Barat. Tingkat keterbelakangan dan pengangguran masih menjadi dasar atas kemiskinan. Kemiskinan setiap wilayah tentunya berbeda-beda. Pemerintah perlu mengetahui wilayah mana yang masuk ke dalam kategori tingkat kemiskinan tinggi dan tingkat kemiskinan rendah sehingga dapat membuat solusi untuk menetapkan skala prioritas dalam memberikan bantuan. Oleh karena itu, dibutuhkan teknik *data mining* yang dapat mengelompokkan tingkat kemiskinan wilayah-wilayah di Jawa Barat yaitu teknik *clustering* dengan algoritma *K-Means*. Tujuan dilakukannya penelitian ini yaitu untuk melakukan pengelompokan pada data kemiskinan di Provinsi Jawa Barat sehingga dapat digunakan sebagai informasi untuk menentukan kebijakan yang tepat untuk menyalurkan bantuan kepada masyarakat dari pemerintahan Jawa Barat. Hasil yang didapatkan berdasarkan pengujian, *cluster* yang diperoleh sebanyak 2 *cluster* dengan *cluster* 0 tingkat kemiskinan tinggi sebanyak 14 wilayah dan *cluster* 1 tingkat kemiskinan rendah sebanyak 13 wilayah. Berdasarkan pengujian, Algoritma *K-Means* memperoleh hasil *Silhouette Coefficient* yaitu 0.576 dan masuk ke dalam kategori *medium structure*. Dengan hasil pengelompokan data kemiskinan tersebut, pemerintah dapat menyalurkan bantuan dengan lebih tepat.

Kata kunci: *Data Mining, K-Means, Kemiskinan*

1. Pendahuluan

Kemiskinan yang cukup serius masih menjadi salah satu masalah, terutama di negara berkembang seperti Indonesia. Kemiskinan pada sebuah daerah dapat menyebabkan terhambatnya pembangunan nasional dalam jangka panjang. Kemiskinan ini merupakan sebuah kondisi dimana kehidupan perekonomian

seseorang atau keluarga serba kekurangan sehingga tidak sanggup untuk memenuhi kebutuhan-kebutuhan pokok yang mencukupi untuk kehidupannya. Kebutuhan pokok yang dimaksudkan adalah minimal memenuhi kebutuhan pangan/makanan, sandang/pakaian, tempat tinggal dan keperluan sosial yang dibutuhkan oleh penduduk ataupun keluarga [1].

Menurut Badan Pusat Statistik (2020) salah satu Provinsi di Indonesia yang mengalami peningkatan kemiskinan yaitu Provinsi Jawa Barat. Pada bulan Maret 2020, tingkat kemiskinan Jawa Barat hingga 3,92 juta jiwa atau 7,88%. Peningkatan tersebut berjumlah sekitar 544,3 ribu jiwa dibanding dengan data sampai bulan September 2019, dimana tingkat kemiskinannya 3,38 juta jiwa atau 6,82% pada saat itu. Hal ini diakibatkan oleh munculnya pandemi covid-19.

Kemiskinan seringkali didasari oleh tingginya tingkat keterbelakangan dan pengangguran. Penanggulangan kemiskinan dapat didukung dengan salah satu strategi seperti adanya data yang akurat. Data tersebut dapat digunakan untuk mengevaluasi tentang kebijakan pemerintah mengenai kemiskinan [2].

Dalam mengatasi dan menanggulangi kemiskinan akibat pandemi covid-19, Pemerintah berupaya memberikan bantuan sosial berupa sembako, subsidi listrik, maupun bantuan untuk UMKM. Dengan adanya beberapa bantuan ini, pemerintah berharap tingkat kemiskinan di Indonesia semakin berkurang [3].

Kemiskinan pada setiap wilayah tentunya berbeda-beda. Pemerintah perlu mengetahui wilayah mana yang masuk kedalam tingkat kemiskinan tinggi dan tingkat kemiskinan rendah berdasarkan kota/kabupaten yang berada di Jawa Barat, sehingga pemerintah dapat dengan cepat mengambil sebuah solusi untuk menetapkan skala prioritas dalam memberikan bantuan kepada masyarakat miskin yang membutuhkan. Oleh sebab itu, diperlukan teknik *data mining* untuk mengelompokkan tingkat kemiskinan wilayah-wilayah di Jawa Barat yaitu dengan teknik *Clustering*. *Data mining* adalah sebuah teknik untuk mengidentifikasi suatu informasi secara menyeluruh terkait pada berbagai *dataset* yang besar untuk mendapat pengetahuan baru. *Data mining* juga dapat dijelaskan sebagai pengolahan untuk memperoleh informasi dengan menggunakan teknik matematika, statistik, *machine learning* dan *artificial intelligence* untuk mengidentifikasi pengetahuan yang bermanfaat serta terikat dari bermacam data *warehouse* atau *database* yang besar [4]. Proses pengelompokkan suatu *record* yang memiliki kesamaan objek kedalam suatu partisi yang berbeda, sehingga data yang dikelompokkan memiliki arti yang bermanfaat disebut *Clustering*. *Clustering* merupakan satu dari beberapa teknik *data mining* yang bersifat *unsupervised learning* dengan tujuan untuk mengelompokkan data yang tidak memiliki label berdasarkan karakteristiknya ke dalam beberapa kelompok [5]. Salah satu algoritma *Clustering* yang terdapat pada *data mining* digunakan untuk memperoleh kelompok yang memiliki data berjumlah banyak menggunakan metode pemisah yang berdasarkan titik dengan waktu perhitungan yang efisien dan cepat merupakan algoritma *K-Means* [6]. *K-Means* juga termasuk kedalam metode untuk

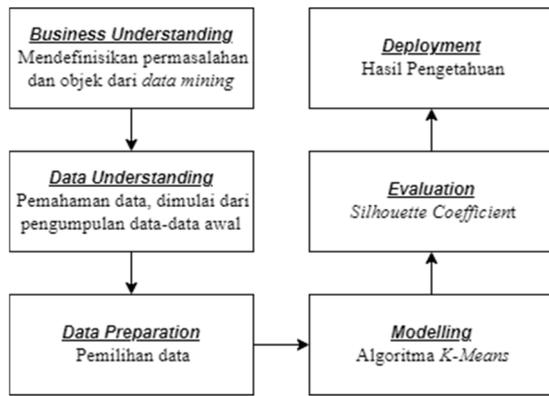
pengelompokan data *non* tingkatan yang berupaya memisahkan data ke dalam suatu kelompok yang sudah ditentukan, sehingga memiliki karakter dan di dalam kelompok yang serupa untuk datanya [7]. Algoritma yang menentukan nilai *Cluster* (*k*) dengan sembarang, dan nilai yang telah ditentukan tersebut menjadi titik pusat dari *Cluster* atau sebagai *centroid* dapat dikatakan sebagai *K-Means* [8].

Penelitian sebelumnya yang menggunakan metode *K-Means* untuk analisis *Clustering*, adalah penelitian [2], penelitian yang dilakukan tersebut memakai algoritma *K-means* dengan pengujian data dengan Microsoft Excel dan *software RapidMiner* untuk mengelompokkan penduduk yang akan menerima bantuan di desa situmekar dengan atribut dusun, umur, dan jumlah keluarga, dengan hasil eksperimen dapat diketahui 3 *Cluster*. Penelitian lainnya yaitu penelitian [9], yang memakai metode *Clustering* untuk mengelompokkan provinsi berdasarkan tingkat kemiskinan menggunakan 5 atribut dengan menggunakan algoritma *K-Means* dan menggunakan *software Weka*, dengan hasil mendapat 3 *Cluster*. Penelitian lainnya yaitu [10] tentang pemanfaatan algoritma *K-Means* untuk pengelompokan data dengan masyarakat miskin berdasarkan provinsi memakai *tools RapidMiner 5.3* dengan perolehan hasil sama dengan perhitungan berdasarkan analisis algoritmanya, dimana diperoleh 8 provinsi masuk kedalam *Cluster* tinggi dan 26 provinsi dengan *Cluster* rendah.

Berdasarkan latar belakang diatas, penelitian ini akan melakukan pengelompokan data kemiskinan di Jawa Barat. Dengan melakukan analisis data kemiskinan, diharapkan pemerintah Provinsi Jawa Barat dapat dengan cepat menentukan kebijakan yang tepat untuk menyalurkan bantuan kepada masyarakat dengan *Cluster* kemiskinan tinggi dan nantinya tingkat kemiskinan di Jawa Barat akan menurun. Berdasarkan dengan nilai akurasi yang baik, maka metode yang digunakan dengan menerapkan teknik *Clustering* dengan memakai algoritma *k-means*. Dalam tahap pemrosesan data pada penelitian ini, menggunakan bahasa pemrograman *Python*.

2. Metode Penelitian

Penelitian ini menggunakan metode *data mining* yaitu CRISP-DM (*Cross-Industry Standard Process for Data Mining*) dimana metode tersebut merupakan salah satu metodologi yang biasa digunakan pada proses *data mining* sebagai rencana dalam memecahkan suatu masalah secara menyeluruh. Metode ini juga telah dilakukan pengembangan oleh analis dari tahun 1996 dari beberapa *industry*, diantaranya Daimler-Benz, ISL, NCR, dan OHRA. Tahapan metodologi CRISP-DM [11] seperti dalam gambar 1 berikut.



Gambar 1. Rancangan Penelitian CRISP-DM

2.1 Business Understanding

Business Understanding merupakan tahap pertama serta bagian paling penting dalam CRISP-DM, tahapan ini mendefinisikan permasalahan, objek dari *data mining* yang akan dilakukan, dan menentukan tujuan penelitian. Serta menyiapkan strategi penyelesaian yang akan dilakukan. *Business understanding* juga dapat didefinisikan sebagai pemahaman dari tujuan penelitian yang membuat perubahan pada pengetahuan dan melakukan pengembangan awal pada *planning* yang telah dilakukan perancangan untuk mencapai suatu maksud tertentu [12].

2.2 Data Understanding

Data Understanding merupakan pemahaman pada data, yang diawali dari mengumpulkan data-data awal, hingga mengidentifikasi kualitas dari data yang telah didapatkan atau untuk mendeteksi data pada bagian yang menarik untuk diasumsikan terhadap informasi yang belum diketahui..

2.3 Data Preparation

Pada tahap ini, meliputi pemilihan data seperti pemilihan tabel, *record*, dan atribut-atribut data yang telah dikumpulkan untuk dapat melakukan pengelompokan dan proses *selection* ke dalam kelompok-kelompok yang telah ditentukan, serta pengolahan data yang nantinya akan diperlukan dalam pemodelan data.

2.4 Modelling

Pada tahap ini dimulai dengan pemilihan teknik serta penggunaan algoritma *data mining* yang diterapkan dalam penelitian. *K-Means* dipilih untuk menjadi metode yang digunakan pada tahapan *modeling*. *K-Means* merupakan algoritma yang menetapkan *K Cluster* dalam partisi set data. Serta algoritma yang sederhana untuk diimplementasikan dan relatif mudah beradaptasi dalam praktiknya [13]. Dalam algoritma *K-Means*, ukuran perbedaan digunakan sebagai pengelompokan objek, perbedaan ini diasumsikan

dalam bentuk jarak, dimana dua objek dikatakan mirip jika jarak keduanya dekat [9].

2.5 Evaluation

Tahap *evaluation* ini, berfokus terhadap model yang telah didapatkan dan diharapkan mempunyai kualitas yang baik. Tahap ini dilakukan keefektifan model sebelum digunakan apakah sudah cocok dengan standar yang ada pada *K-Means Clustering* dan pada tahapan awal tidak ada yang dilewati hingga tahap pemodelan berakhir [14]. *Silhouette Coefficient* digunakan pada tahap evaluasi ini dengan fungsi untuk menguji kualitas dari *cluster* yang telah didapatkan. Selanjutnya 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 *Silhouette Coefficient* dekat dengan rentang nilai 1 dan sebaliknya, *Cluster* yang buruk memiliki nilai dari rata-rata yang dihasilkan *Silhouette Coefficient* dekat dengan rentang nilai -1. Rentang nilai ini dipergunakan untuk menunjukkan korelasi kemiripan dari data yang telah dikelompokkan ke dalam *Cluster*. Berikut pada tabel 1 merupakan kriteria dalam mengukur nilai *Silhouette Coefficient* [15].

Tabel 1. Kriteria Silhouette Coefficient

<i>Silhouette Coefficient</i>	Kriteria Penilaian
$0.7 < Silhouette \leq 1.0$	<i>Strong Structure</i>
$0.5 < Silhouette \leq 0.7$	<i>Medium Structure</i>
$0.25 < Silhouette \leq 0.5$	<i>Weak Structure</i>
$Silhouette \leq 0.25$	<i>No Structure</i>

2.6 Deployment

Setelah tahap *evaluation* yang berdasarkan hasil pengujian, maka dilakukan penyusunan untuk menjadi sebuah pengetahuan. Dengan pengimplemetasian dari keseluruhan model yang telah dirancang. Tahapan ini melibatkan pengguna karena untuk mengerti bagaimana langkah yang akan dilakukan untuk menggunakan model yang telah dibuat.

3. Hasil dan Pembahasan

Berdasarkan hasil dari penelitian, pengelompokan (*clustering*) dengan tahapan pada metode CRISP-DM serta penggunaan algoritma *K-Means* dan menggunakan bahasa pemrograman *Python* untuk penelitiannya, yang dibantu dengan *tools Google Collab*. Dengan pembahasan sebagai berikut.

3.1 Business Understanding

Pada tahap ini, dibutuhkan pemahaman mengenai pentingnya suatu data kemiskinan untuk melakukan pengelompokan kabupaten/kota dengan tingkat kemiskinan yang berbeda berdasarkan status bekerja pada kabupaten/kota di Provinsi Jawa Barat. Tujuan dari penelitian ini, yaitu dengan menggunakan *data mining*, pemerintah Jawa Barat dapat mengetahui kelompok kemiskinan tiap kabupaten/kotanya yang termasuk ke dalam kemiskinan tinggi dan rendah, berdasarkan status bekerjanya. Dengan melakukan pengelompokan (*clustering*), pemerintah dapat melakukan kebijakan untuk penyaluran bantuan kemiskinan.

3.2 Data Understanding

Penelitian ini memakai data publik yang diperoleh dari publikasi kemiskinan kabupaten/kota di Provinsi Jawa Barat pada tahun 2015-2020 yang dikeluarkan oleh Badan Pusat Statistik Provinsi Jawa Barat. Informasi pada publikasi ini berisikan data tentang kemiskinan pada tingkatan kabupaten/kota di Jawa Barat serta berisi indikator/atribut yang terikat dengan kemiskinan.

Data yang digunakan sebanyak 27 data pada tahun 2020 dengan menggunakan format *.csv (comma separated values)*. Data tersebut berisi 4 atribut yang akan digunakan dalam penelitian yaitu wilayah jawa barat, tidak bekerja, bekerja di sektor pertanian, dan bekerja bukan di sektor pertanian. Dapat dilihat gambar 2.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27 entries, 0 to 26
Data columns (total 4 columns):
#   Column                               Non-Null Count  Dtype
---  ---                               -----
0   Wilayah Jawa Barat                   27 non-null     object
1   Tidak Bekerja                        27 non-null     float64
2   Bekerja di Sektor Pertanian          27 non-null     float64
3   Bekerja Bukan di Sektor Pertanian    27 non-null     float64
dtypes: float64(3), object(1)
memory usage: 992.0+ bytes
```

Gambar 2. Informasi Data

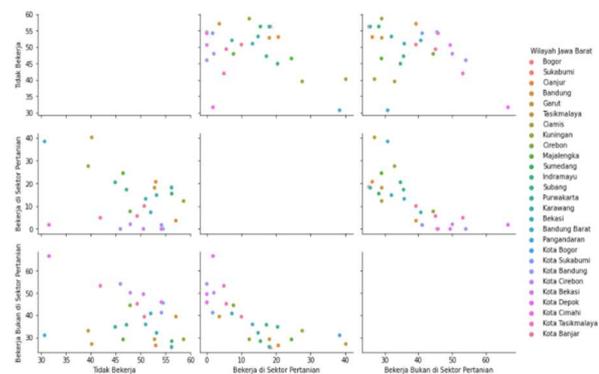
3.3 Data Preparation

Pada proses *data preparation*, langkah pertama yaitu pemilihan atribut. Atribut tersebut digunakan untuk proses pada tahap *modeling*. Tahap ini memilih semua atribut dikarenakan semua atribut saling berhubungan dan berpengaruh terhadap hasil yang akan ditampilkan dari proses *modelling*. Selanjutnya menghilangkan *missing value* yang berada pada atribut yang memiliki data *null* atau kosong. Gambar 3 menunjukkan bahwa data yang digunakan tidak ada nilai kosong atau *missing value*. Jadi, data tidak ada yang dihilangkan.

```
data.isnull().sum()
Wilayah Jawa Barat      0
Tidak Bekerja           0
Bekerja di Sektor Pertanian  0
Bekerja Bukan di Sektor Pertanian  0
dtype: int64
```

Gambar 3. Mencari *missing value*

Tahap selanjutnya yaitu transformasi data. Pertama, dilakukan dengan pemilihan fitur yang dapat dilakukan *cluster* dilihat dari visualisasi datanya. Dimana, data awal divisualisasikan lalu dilihat sebaran datanya, yaitu dengan cara memilih visualisasi yang menyerupai *Cluster* untuk dijadikan fitur. Persebaran data dapat dilihat pada gambar 4.



Gambar 4. Visualisasi Persebaran Data

Pada gambar 4 tersebut dapat dilihat persebaran data yang berkaitan antar atributnya. Dengan atribut wilayah Jawa Barat sebagai titik-titik persebaran datanya, dan atribut lain yaitu tidak bekerja, bekerja di sektor pertanian dan bekerja bukan di sektor pertanian menggambarkan keterkaitan dengan wilayah Jawa Barat yaitu sebagai fitur. Visualisasi tersebut digunakan sebagai penentu dalam memilih fitur yang cocok digunakan untuk proses *modelling*. Dimana, berdasarkan gambar 4 tersebut fitur yang dipilih untuk di *Clustering* yaitu bekerja di sektor pertanian dan bekerja bukan di sektor pertanian. Karena dalam visualisasinya, data antar fitur tersebut lebih terlihat *clustering*nya.

Tahap kedua dalam transformasi data dengan cara normalisasi data atau bisa disebut standarisasi ukuran datanya dengan menggunakan fungsi *Min-Max Scale*. Data yang dinormalisasi merupakan data hasil pemilihan fitur. Hasil normalisasi terdapat pada tabel 2.

Tabel 2. Hasil Normalisasi Data

No	Bekerja di Sektor Pertanian	Bekerja Bukan di Sektor Pertanian
0	0.24881811	0.33705141
1	0.45807415	0
2	0.51405822	0.02303589
...
24	0	0.49587779
25	0.12241851	0.6741028
26	0.13958696	0.47793404

Cluster optimal tersebut diperoleh dari perhitungan *K* dengan memilih jarak inerti antara satu dengan lainnya jauh. Selanjutnya mengelompokkan data sesuai dengan *Clusternya*. Jarak terdekat antara data dengan titik pusat (*centroid*) dari suatu *Cluster* akan menentukan kelompok *Cluster* dari data tersebut. Titik *centroid* tiap *Cluster* ditunjukkan pada Tabel 3.

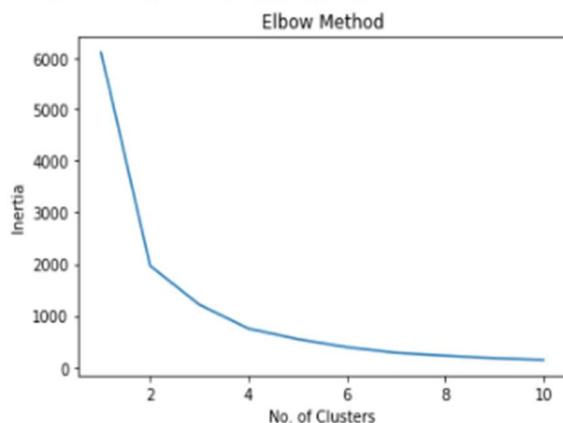
Tabel 3. Titik Centroid *Cluster*

<i>Cluster</i>	Tidak Bekerja	Bekerja di Sektor Pertanian	Bekerja Bukan di Sektor Pertanian
0	49.01	21.37	30.13
1	49.06	3.43	47.27

3.4 Modelling

Selanjutnya tahap *modelling*. Langkah awal yang harus dilakukan yaitu dengan menentukan jumlah *Cluster* yang nantinya akan digunakan untuk proses pengelompokkan menggunakan *K-Means*. Proses penentuan jumlah *Cluster* dilakukan dengan memakai metode *Elbow*. Dapat dilihat pada gambar 5 yaitu penentuan *Cluster*..

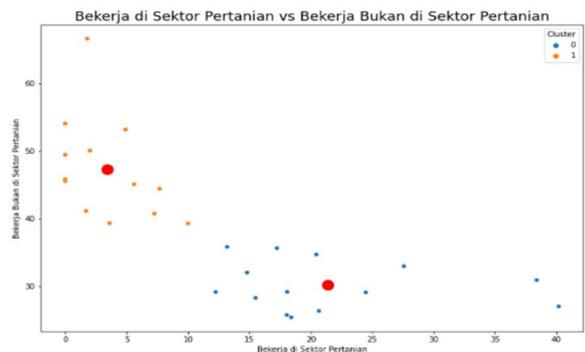
```
K : 1 Inertia : 6119.273533333333
K : 2 Inertia : 1968.2701978021973
K : 3 Inertia : 1216.1050442307692
K : 4 Inertia : 751.7668271212119
K : 5 Inertia : 547.6430621212119
K : 6 Inertia : 391.8627914285713
K : 7 Inertia : 282.8443199999999
K : 8 Inertia : 226.49111666666664
K : 9 Inertia : 179.25856499999995
K : 10 Inertia : 144.98465833333333
```



Gambar 5. Grafik *Cluster* Optimal dengan Metode *Elbow*

Berdasarkan pada hasil metode *elbow* yang dapat dilihat pada gambar 5, didapat kesimpulan bahwa dari data yang telah diolah menghasilkan 2 *Cluster* optimal.

Dilihat dari tabel 3 maka dapat disimpulkan bahwa *Cluster* 0 merupakan *Cluster* dengan tingkat kemiskinan yang tinggi, dan *Cluster* 1 dengan tingkat kemiskinan yang rendah. Hasil pengelompokkan tersebut yaitu 14 wilayah Jawa Barat pada *cluster* 0, dan 13 wilayah Jawa Barat masuk ke *Cluster* 1. Selanjutnya visualisasi data dengan *Cluster* yang telah ditentukan dengan menggunakan algoritma *K-Means* terdapat pada gambar 6 berikut.



Gambar 6. Visualisasi Data *Cluster*

Berdasarkan visualisasi pengelompokkan data dengan 2 *Cluster*, dapat dilihat pada gambar 6 bahwa persebaran data yang berada dalam *Cluster* 0 lebih banyak mengarah pada bekerja di sektor pertanian, dan data yang masuk *Cluster* 1 lebih banyak mengarah pada bekerja bukan di sektor pertanian. Visualisasi tersebut menggunakan beberapa warna untuk membedakan *Cluster*, dengan titik pusat berwarna merah, warna biru yaitu untuk *Cluster* 0, dan warna oren untuk *Cluster*.

3.5 Evaluation

Pada tahap *evaluation*, digunakan teknik *Silhouette Coefficient* dimana bertujuan untuk menguji kualitas *Cluster* dari data yang telah didapatkan sebelumnya

pada tahap *modelling*. Maka, didapatkan nilai rata-rata *Silhouette* yang terdapat pada gambar 7.

Nilai Rata - Rata *Silhouette* : 0.5765876427369332

Gambar 7. Nilai Rata-rata *Silhouette*

Didapatkan nilai *silhouette* dari proses *K-Means* sebesar 0.576. Berdasarkan tabel I kriteria *Silhouette Coefficient*, maka termasuk dalam kriteria *Medium Structure*. Nilai *silhouette* sudah dapat dikatakan baik jika nilainya > 0.5, jadi untuk pengujian kualitas *Cluster* dengan pengelompokan kemiskinan menggunakan algoritma *K-Means* ini memiliki kualitas yang baik.

3.6 Deployment

Berdasarkan dengan hasil pengolahan data pada tahap sebelumnya yang telah dilakukan, tahap *deployment* dilakukan penyusunan dari pengetahuan yang dihasilkan dari menggunakan data kemiskinan untuk pengelompokan dengan algoritma *K-Means* sehingga dapat menjadi suatu informasi yang bermanfaat bagi pemerintah. Tabel 4 adalah hasil *Clustering K-Means* dengan masing-masing anggota dari *Cluster* yang terbentuk..

Tabel 4. Hasil *Clustering K-Means*

Cluster	Hasil <i>Clustering K-Means</i> (Kabupaten/Kota)
0	Sukabumi, Cianjur, Garut, Tasikmalaya, Ciamis, Kuningan, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bandung Barat, Pangandaran.
1	Bogor, Bandung, Cirebon, Bekasi, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar.

Berdasarkan hasil *cluster K-Means* tersebut, dari total 27 data wilayah yang menunjukkan kabupaten/kota di Jawa Barat, wilayah tersebut dikelompokkan berdasarkan *clusternya*. Yang masuk ke dalam cluster 0 sebanyak 14 wilayah kabupaten/kota di Jawa Barat. Sedangkan untuk cluster 1, yang masuk sebanyak 13 wilayah kabupaten/kota di Jawa Barat. Dengan nama-nama kabupaten dan kota berdasarkan kelompoknya yang dapat dilihat pada tabel 4.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dengan memakai metode *K-Means Clustering* dalam melakukan pengelompokan data kemiskinan di Jawa Barat dengan menggunakan *tools Google Collab* dan

bahasa pemrograman *Python*. Pengelompokan menggunakan 4 atribut yaitu wilayah Jawa Barat, tidak bekerja, bekerja di sektor pertanian, dan bekerja bukan di sektor pertanian. Dimana, dalam pengujian tersebut menggunakan nilai $K = 2$ atau 2 *Cluster* yang didapat dari metode *elbow*. Dari hasil perhitungan menggunakan metode *K-Means*, didapatkan informasi yang baru berupa tingkat kemiskinan yang terbagi menjadi 2 *Cluster* yaitu *Cluster 0* dengan tingkat kemiskinan tinggi sebanyak 14 wilayah dan *cluster 1* dengan tingkat kemiskinan rendah sebanyak 13 wilayah.. Diharapkan dengan adanya informasi tersebut, pemerintah Jawa Barat dapat dengan cepat menentukan kebijakan yang tepat untuk menyalurkan bantuan kepada masyarakat yang membutuhkan. Hasil evaluasi dari penggunaan algoritma *K-Means* yang dilakukan untuk pengelompokan dari data kemiskinan di Provinsi Jawa Barat menggunakan metode *Silhouette Coefficient* dengan menghasilkan nilai sebesar 0.576, dari nilai tersebut dapat disimpulkan bahwa hasil *Cluster* yang telah didapatkan cukup baik serta masuk kategori *medium structure*.

Daftar Rujukan

- [1] N. I. Febianto and N. Palasara, "Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 2, pp. 130–140, 2019, doi: 10.32736/sisfokom.v8i2.653.
- [2] F. Andrianti, R. Firmansyah, U. Adhirajasa, R. Sanjaya, A. R. Sanjaya, and D. Mining, "Penerapan Clustering Data Kurang Mampu Di Desa Situmekar Menggunakan Algoritma K-," vol. 1, no. 1, pp. 88–95, 2020.
- [3] C. W. Prasetyandari, "Dampak Covid-19 Bagi Roda Perekonomian Bagi Masyarakat Indonesia," *J. Imagine*, vol. 1, no. 1, pp. 12–16, 2021, doi: 10.35886/imagine.v1i1.172.
- [4] M. R. Muttaqin and M. Defriani, "Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 121–129, 2020, doi: 10.33096/ilkom.v12i2.542.121-129.
- [5] T. M. A. K-means, "Isbn : 2685-5852 fmipa unimus 2019 pemetaan kemiskinan kabupaten/kota di provinsi jawa tengah menggunakan algoritma k-means," pp. 553–557, 2019.
- [6] E. Fammaldo and L. Hakim, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar," *J. Ilm. Teknol. Infomasi Terap.*, vol. 5, no. 1, pp. 23–31, 2019, doi: 10.33197/jitter.vol5.iss1.2018.249.
- [7] R. Hidayat, R. Wasono, and M. Y. Darsyah, "Pengelompokan Kabupaten/Kota Di Jawa Tengah Menggunakan Metode K-Means Dan Fuzzy C-Means," *Pros. Semin. Nas. Int.*, no., pp. 240–250, 2017, [Online]. Available: <https://jurnal.unimus.ac.id/index.php/psn12012010/article/view/3017/2932>.
- [8] Purnia Silvi Dini, "Indonesian Journal of Computer Science," *STMIK Indones. Padang*, vol. 6, no. 1, p. 62, 2020.
- [9] A. Bahaudin, A. Fatmawati, and F. Permata Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *J. Manaj. Inform. dan Sist. Inf.*, vol. 4, no. 1, p. 1, 2021, doi: 10.36595/misi.v4i1.216.
- [10] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 76–83, 2020, doi: 10.47065/bits.v2i2.492.
- [11] A. R. Wibowo and A. Jananto, "Implementasi Data Mining Metode Asosiasi Algoritma FP-Growth Pada Perusahaan

- Ritel,” *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, p. 200, 2020, doi: 10.35585/inspir.v10i2.2585.
- [12] D. Dharmayanti, A. M. Bachtiar, and A. C. Prasetyo, “Penerapan Metode Clustering Untuk Membentuk Kelompok Belajar Menggunakan Di Smpn 19 Bandung,” *Komputa J. Ilm. Komput. dan Inform.*, vol. 6, no. 2, pp. 49–56, 2017, doi: 10.34010/komputa.v6i2.2477.
- [13] Yuni Radana Sembiring, Saifullah, and Riki Winanjaya, “Implementasi Data Mining Dalam Mengelompokkan Jumlah Penduduk Miskin Berdasarkan Provinsi Menggunakan Algoritma,” *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen) Vol. 2, No. 2*, vol. 2, no. 2, pp. 125–132, 2021.
- [14] N. Mirantika, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat,” *Nuansa Inform.*, vol. 15, no. 2, pp. 92–98, 2021, doi: 10.25134/nuansa.v15i2.4321.
- [15] R. A. Farissa, R. Mayasari, and Y. Umidah, “Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokkan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsembung,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 109–116, 2021, doi: 10.30871/jaic.v5i1.3237.