

Clustering Data Penduduk Miskin Dampak Covid-19 Menggunakan Algoritma *K-Medoids*

Novi Widiawati ^{1*}, Betha Nurina Sari ^{2**}, Tesa Nur Padilah ^{3*}

* Teknik Informatika, Universitas Singaperbangsa Karawang

novi.widiawati17164@student.unsika.ac.id¹, betha.nurina@staff.unsika.ac.id², tesa.nurpadilah@staff.unsika.ac.id³

Article Info

Article history:

Received 2021-08-14

Revised 2022-05-11

Accepted 2022-05-12

Keyword:

Clustering,
Data Mining,
K-Medoids,
Penduduk Miskin,
Silhouette Coefficient.

ABSTRACT

Kemiskinan merupakan masalah yang mendasar, kemiskinan bisa berakibat pada terhambatnya pembangunan nasional. Ada beberapa aspek yang berkaitan dengan kemiskinan yaitu faktor ekonomi, politik, dan psikososial. Secara ekonomi, kemiskinan diartikan sebagai kurangnya sumber daya untuk memenuhi kebutuhan hidup dan meningkatkan kesejahteraan. Pada penelitian ini data yang digunakan pada tahun 2020 yang bersumber dari Badan Pusat Statistika. Dalam upaya menemukan kasus kemiskinan dampak covid-19 dapat menggunakan *Data Mining*. Tujuan dari penelitian ini untuk pengelompokan atau pemetaan daerah di Indonesia berdasarkan tingkat kemiskinan dampak covid-19 dapat dilakukan menggunakan algoritma *k-medoids*. Penelitian yang akan dilakukan dengan langkah *data mining* yaitu CRISP-DM terdiri dari 6 fase yaitu pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), pengolahan data (*data preparation*), pemodelan (*modelling*), evaluasi (*evaluation*), dan penyebaran (*deployment*). Algoritme yang digunakan pada penelitian ini yaitu *K-Medoids*. Pengukuran menggunakan bahasa *R* dengan bantuan fungsi Pamk sehingga hasil yang didapatkan pada *dataset* Penduduk Miskin Tahun 2020 memiliki *cluster* optimal sebanyak 2 *cluster*. *Cluster1* dengan jumlah 121 kabupaten/kota dengan kategori tinggi, sedangkan *cluster2* dengan jumlah 427 dengan kategori rendah. Hasil dari evaluasi nilai *Silhouette Coefficient* sebesar 0,4735719.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Kemiskinan merupakan masalah yang sangat mendasar dan telah menjadi fokus pemerintah di negara mana pun. Indonesia terkategori negara berkembang, tetapi kemiskinan masih menjadi isu yang wajib dicermati. Kemiskinan bisa berakibat pada terhambatnya pembangunan nasional [1]. Keadaan yang dimana kehidupan serba kekurangan sehingga tidak mampu memenuhi kebutuhan hidup yang sesuai. Ada beberapa aspek yang berkaitan dengan kemiskinan yaitu faktor ekonomi, politik, dan psikososial. Secara ekonomi, kemiskinan diartikan sebagai kurangnya sumber daya untuk memenuhi kebutuhan hidup dan meningkatkan kesejahteraan. Peningkatan jumlah penduduk di perkotaan telah menimbulkan berbagai masalah seperti terjadinya kejahatan, kemacetan lalu lintas, pengangguran, kesulitan dalam

penyediaan tempat tinggal yang layak dan fasilitas sosial lainnya [2].

Badan Pusat Statistik (2020) mencatat antara tahun 2010 hingga September 2020, angka kemiskinan di Indonesia menurun baik dalam jumlah ataupun persentase kecuali pada bulan September 2013, Maret 2020 serta September 2020. Peningkatan harga kebutuhan pokok memicu peningkatan harga BBM pada September 2013 serta Maret 2015. Pada masa yang sama, jumlah serta persentase penduduk miskin bertambah di bulan Maret 2020 serta September 2020 diakibatkan oleh pandemi covid-19 yang menyerang Indonesia. Gambar 1 menampilkan angka kemiskinan dari 2010 hingga September 2020.



Gambar 1 Jumlah dan Persentase Penduduk Miskin

Maret 2010 - September 2020

Pada September 2020, jumlah penduduk miskin di Indonesia mencapai 27,55 juta jiwa. Dibandingkan dengan Maret 2020, jumlah penduduk miskin meningkat 1,13 juta jiwa. Sementara jika dibandingkan dengan September 2019, jumlah penduduk miskin meningkat 2,76 juta jiwa. Persentase penduduk miskin pada September 2020 tercatat sebesar 10,19 persen, meningkat 0,41 persen poin terhadap Maret 2020 dan meningkat 0,97 persen poin terhadap September 2019. Menurut wilayah tempat tinggal, antara Maret 2020 dan September 2020, penduduk miskin perkotaan meningkat 876,5 ribu, sedangkan pedesaan meningkat 249,1 ribu. Angka kemiskinan di perkotaan meningkat dari 7,38% menjadi 7,88%. Sementara di pedesaan, proporsi ini meningkat dari 12,82% menjadi 13,20%.

TABEL I
JUMLAH DAN PERSENTASE PENDUDUK MISKIN MENURUT DAERAH
SEPTEMBER 2019-SEPTEMBER 2020

SUMBER : BRS SUSENAS (2020)

Daerah/Tahun	Jumlah Penduduk Miskin (Juta Orang)	Persentase Penduduk Miskin
(1)	(2)	(3)
Perkotaan		
September 2020	9,86	6,56
Maret 2020	11,16	7,38
September 2020	12,04	7,88
Pedesaan		
September 2019	14,93	12,60
Maret 2020	15,26	12,82
September 2020	15,51	13,20
Total		
September 2019	24,79	9,11
Maret 2020	26,42	9,78
September 2020	27,55	10,19

Berdasarkan Tabel 1 dapat diketahui bahwa tingkat kasus kemiskinan terus meningkat pada setiap tahunnya. Sumber daya yang kurang akan mempengaruhi kesejahteraan dan

kebutuhan hidup, karena kemiskinan yang tidak adanya pengembangan dalam mengatasinya. Salah satu cara untuk membantu mengatasi kasus kemiskinan yaitu mencari daerah dengan tingkat kemiskinan tertinggi. Tindakan penanganan dapat menjadi lebih efektif sehingga lebih berfokus untuk daerah kemiskinan yang terkategori tinggi. *Data mining* merupakan teknik yang digunakan untuk mengetahui prioritas dengan kategori daerah kemiskinan tertinggi. Oleh karena itu, *data mining* memegang peranan yang sangat berarti dalam bermacam bidang kehidupan (antara lain industri, keuangan, cuaca, serta iptek). Tujuan dari *data mining* untuk menciptakan tren ataupun pola yang diperlukan dalam *dataset* besar untuk menentukan nilai. *Clustering* adalah metode yang terdapat pada *data mining* untuk pengelompokan data [8-12], tujuan dari *clustering* adalah untuk mengelompokkan *cluster* dari objek, titik, dokumen dan pola. Perbandingan antara objek kelompok *cluster* lain dengan pengelompokan objek *cluster* yang memiliki persamaan antara kelompok lain [5].

Pada penelitian yang dilaksanakan [6] dalam pengelompokan data sebaran anak cacat lebih baik menggunakan algoritme *k-medoids* karena hasil validasi lebih baik, dibandingkan dengan menggunakan algoritme *k-means*. Selain itu pada penelitian lain [3] menyatakan algoritme *k-means* memiliki kelemahan yang *sensitive* terhadap data *noise* dan *outlier*, sehingga untuk mengatasi kelemahan tersebut dengan menggunakan algoritme *k-medoids*. Dalam menangani *dataset* yang kecil algoritme *k-medoids* cukup efisien dibandingkan algoritme *k-means* [4]. Pada algoritma *k-medoids* sebuah *cluster* yang disebut dengan *medoids* merupakan objek yang terpilih. Tujuan dari *k-medoids* yaitu mengurangi sensitivitas yang dihasilkan dari nilai ekstrim yang terdapat pada *dataset*, pengamatan mean tidak didasarkan pada *medoids* untuk setiap *cluster* [13-25].

Berdasarkan penelitian dan masalah di atas maka akan dilakukan penelitian yang berjudul “*Clustering Data Penduduk Miskin Dampak Covid-19 Menggunakan Algoritme K-Medoids*”. Pada penelitian ini akan dilakukan *clustering* algoritme *k-medoids* dalam pendekatan *data mining* yang bertujuan untuk mengetahui wilayah dengan tingkat kemiskinan tertinggi.

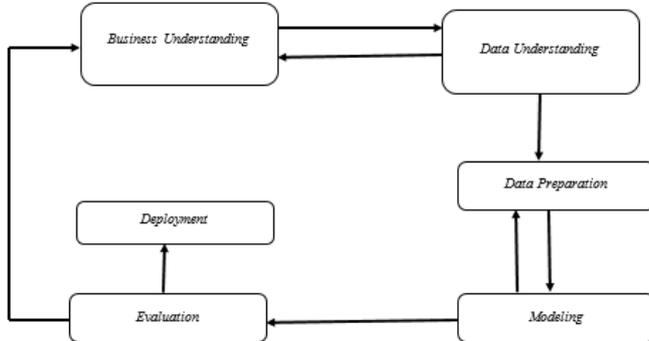
II. METODE PENELITIAN

A. Metodologi Penelitian

Pada penelitian ini metodologi yang digunakan yaitu CRISP-DM [13] dengan menggunakan metode analisis deskriptif dan metode penelitian pendekatan kuantitatif, artinya penelitian yang dilakukan adalah menekankan hasil analisis pada basis data yang bertujuan untuk mendapatkan gambaran yang jelas untuk mengelompokkan wilayah berdasarkan data kemiskinan akibat covid-19 di Indonesia. Metodologi penelitian CRISPM-DM (*Cross-Industry Standard Process For Data Mining*) yang dilakukan terdiri dari 6 fase yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*.

B. Rancangan Penelitian

Berikut alur penelitian clustering data penduduk miskin dampak covid-19 dimana langkah-langkah yang diambil berdasarkan metode CRISP-DM (*Cross Industry Standar Proses for Data Mining*) dapat dilihat pada Gambar 2 sebagai berikut.



Gambar 2. Rancangan Penelitian

III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian

Hasil penelitian ini adalah data kemiskinan dampak covid-19 pada tahun 2020 yang dari <https://www.bps.go.id/> yang telah dikelompokkan sesuai dengan kelasnya. Algoritma yang digunakan dalam *clustering* data adalah *k-medoids*, kemudian hasil pengolahan data kemiskinan akan divisualisasikan ke dalam bentuk *cluster* yang dimana pada data tersebut terbagi menjadi 2 kategori yaitu tinggi dan rendah.

1) *Business Understanding*: Tahapan ini meliputi studi literatur mengenai teori-teori yang relevan dengan *data mining*, algoritme *k-medoids*, juga teknik *clustering* data kemiskinan. Permasalahan yang dianalisis peneliti pada penelitian ini yaitu terkait *clustering* wilayah berdasarkan data kemiskinan dampak covid-19 di Indonesia.

2) *Data understanding*: *Data understanding* atau pemahaman data merupakan tahapan untuk mengumpulkan data awal dan mempelajari data tersebut. Adapun proses pemahaman data tersebut dilakukan agar mengetahui struktur data dan penanganan terhadap data.

3) *Data Preparation*: Tahap *pre-processing* atau persiapan data ini merupakan tahap dimana data yang sudah diperoleh dilakukan persiapan dan pembersihan data. Data kemiskinan yang telah dikumpulkan sebelumnya, masih berupa data yang tidak terstruktur. Oleh karena itu, diperlukan tahap *pre-processing* untuk mengubahnya menjadi data yang lebih terstruktur. Pada penelitian ini *pre-processing* terdiri dari *select data*, *data cleaning*, *data integration*, dan *format data*. Berikut ini merupakan proses dan hasil dari *pre-processing* data tersebut.

a. Select Data

Penelitian ini menggunakan *dataset* kemiskinan dampak covid-19 kabupaten/kota di Indonesia pada tahun 2020 dengan atribut nama wilayah, persentase penduduk miskin, indeks kedalaman kemiskinan, indeks keparahan kemiskinan,

indeks pembangunan manusia, dan jumlah penduduk miskin. *Dataset* diperoleh dari <https://www.bps.go.id/>. Berikut merupakan tabel kasus kemiskinan yang ada di kabupaten/kota di Indonesia.

TABEL 2
DATASET PENDUDUK MISKIN TAHUN 2020

	Nama Wilayah	Persentase Penduduk Miskin	Indeks Kedalaman Kemiskinan	Indeks Keparahan Kemiskinan	Indeks Pembangunan Manusia	Jumlah Penduduk Miskin
1						
2	ACEH	14,99	2,72	0,71	71,99	814,31
3	Simeulue	18,49	2,21	0,47	66,03	17,34
4	Aceh Singkil	20,2	3,55	1,03	68,94	25,43
5	Aceh Selatan	12,87	2,13	0,46	67,12	30,91
6	Aceh Tenggara	13,21	1,62	0,36	69,37	28,98
7	Aceh Timur	14,08	2,29	0,61	67,63	62,34
8	Aceh Tengah	15,08	2,89	0,95	73,24	32,48
9	Aceh Barat	18,34	2,74	0,68	71,38	39,06
10	Aceh Besar	13,84	2,4	0,61	73,56	59,7
11	Pidie	19,23	2,2	0,45	70,63	86,39
536	Sami	13,87	1,56	0,29	63,63	5,7
537	Keerom	16,32	2,79	0,7	66,4	9,42
538	Waropen	29,54	9,9	4,12	64,94	9,44
539	Supiori	36,91	10,52	3,8	62,3	7,78
540	Mamberamo Raya	28,38	4,12	1,01	51,78	6,98
541	Nduga*	36,72	7,57	2,25	31,55	36,54
542	Lanny Jaya	38,13	13,87	6,51	47,86	68,62
543	Mamberamo Tengah	36,41	9,11	2,84	47,57	17,72
544	Yalimo	32,82	11,47	4,86	48,34	20,84
545	Puncak	36,96	8,61	2,63	43,04	42,43
546	Dogiyai	28,62	5,55	1,86	54,84	28,31
547	Intan Jaya	40,71	8	2,17	47,73	20,46
548	Deiyai	41,76	11,44	3,9	49,46	30,98
549	Kota Jayapura	11,16	1,98	0,55	79,94	33,8

Pada Tabel 2 kasus kemiskinan di Indonesia Tahun 2020 terdiri dari 6 variabel dan 548 objek data. Dari *dataset* di atas dilakukan proses *select data* atau pemilihan data, dimana hanya data-data tertentu yang akan digunakan pada proses *data mining*. Berikut merupakan data hasil seleksi *dataset* penduduk miskin tahun 2020 dapat dilihat pada Tabel 3.

TABEL 3
DATASET HASIL SELEKSI PENDUDUK MISKIN TAHUN 2020

Atribut	Tipe Data	Keterangan
Nama Wilayah	Character	Nama Wilayah
Persentase Penduduk Miskin	Numeric	Persentase Penduduk Miskin
Indeks Kedalaman Kemiskinan	Numeric	Indeks Kedalaman Kemiskinan
Indeks Keparahan Kemiskinan	Numeric	Indeks Keparahan Kemiskinan
Indeks Pembangunan Manusia	Numeric	Indeks Pembangunan Manusia
Jumlah Penduduk Miskin	Numeric	Jumlah Penduduk Miskin

b. Data Cleaning

Setelah pemilihan data dilakukan, nilai-nilai yang hilang diproses, duplikat data dan data yang bertentangan atau tidak konsisten. Pada *dataset* penduduk miskin tahun 2020 dilakukan *data pre-processing*. Gambar 3 merupakan hasil *summary* dari *dataset* yang digunakan, pada hasil *summary* tersebut ditemukan nilai NA atau *missing value*.

```

> summary(Data_Skrripsi)
Nama wilayah      Persentase Penduduk Miskin Indeks Kedalaman Kemiskinan
Length:548      Min. : 2.02      Min. : 0.070
Class :character 1st Qu.: 6.78      1st Qu.: 0.870
Mode :character  Median :10.05     Median : 1.405
                Mean  :11.86      Mean  : 2.004
                3rd Qu.:14.55    3rd Qu.: 2.345
                Max.  :41.76      Max.  :13.870

Indeks Keparahan Kemiskinan Indeks Pembangunan Manusia Jumlah Penduduk Miskin
Min. :0.0400      Min. :31.55      Min. : 1.36
1st Qu.:0.1750    1st Qu.:66.51    1st Qu.: 15.70
Median :0.3100    Median :69.39    Median : 30.14
Mean  :0.5397     Mean  :69.72     Mean  : 66.47
3rd Qu.:0.5850    3rd Qu.:72.73    3rd Qu.: 77.33
Max.  :6.9900     Max. :86.61      Max. :911.37
NA's  :1          NA's  :7
    
```

Gambar 3. Hasil Summary Dataset Sebelum Dilakukan Pre-Processing

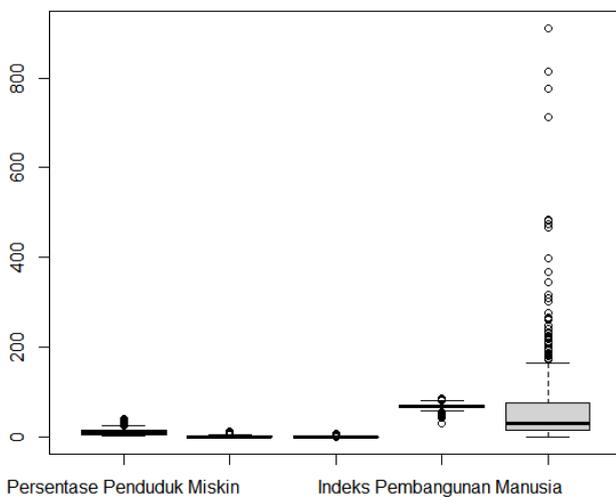
Dalam mengatasi missing value dapat dilakukan dengan mengubah nilai tersebut dengan metode mean. Gambar 4 merupakan hasil summary dari dataset yang digunakan, pada hasil summary tersebut tidak ditemukan nilai NA atau missing value, maka tahap penelitian dilanjutkan ke tahap berikutnya.

```

> summary(Data_Skrripsi)
Nama wilayah      Persentase Penduduk Miskin Indeks Kedalaman Kemiskinan
Length:548      Min. : 2.02      Min. : 0.070
Class :character 1st Qu.: 6.78      1st Qu.: 0.870
Mode :character  Median :10.05     Median : 1.405
                Mean  :11.86      Mean  : 2.004
                3rd Qu.:14.55    3rd Qu.: 2.345
                Max.  :41.76      Max.  :13.870

Indeks Keparahan kemiskinan Indeks Pembangunan Manusia Jumlah Penduduk Miskin
Min. :0.0400      Min. :31.55      Min. : 1.36
1st Qu.:0.1775    1st Qu.:66.51    1st Qu.: 15.88
Median :0.3150    Median :69.39    Median : 30.66
Mean  :0.5397     Mean  :69.72     Mean  : 66.47
3rd Qu.:0.5825    3rd Qu.:72.73    3rd Qu.: 76.84
Max.  :6.9900     Max. :86.61      Max. :911.37
    
```

Gambar 4. Hasil Summary Dataset Setelah Dilakukan Pre-Processing



Gambar 5. Hasil Bloxpot Dataset

Berdasarkan Gambar 5 menunjukkan hasil boxplot dataset di atas dapat dilihat bahwa terdapat data outlier pada variabel jumlah penduduk miskin yang memiliki jarak atau nilai yang cukup jauh dari data lain, sehingga pengelompokkan lebih cocok menggunakan metode k-medoids agar mendapatkan hasil yang lebih representatif.

c. Data Integration

Pada tahap ini data integrasi adalah melakukan penggabungan semua dataset yang dibutuhkan ke dalam satu tabel. Pada penelitian ini hanya menggunakan satu tabel, maka tahap ini tidak dilakukan.

d. Format Data

Pada tahap ini data yang masih menjadi kategorik ditransformasikan ke dalam bentuk numerik, dalam proses perubahan bentuk data dari kategorik ke numerik adalah dengan memberikan angka yang berbeda ke setiap kategorik yang untuk variabel pada dataset. Kemudian dilakukan standarisasi data dengan menggunakan fungsi scale, pada proses normalisasi menggunakan bantuan tools Rstudio sehingga didapatkan hasil seperti pada Tabel 4. Berikut merupakan hasil normalisasi Tabel 4 pada dataset penduduk miskin tahun 2020.

TABEL 4
HASIL NORMALISASI DATASET PENDUDUK MISKIN TAHUN 2020

	Persentase Penduduk Miskin	Indeks Kedalaman Kemiskinan	Indeks Keparahan Kemiskinan	Indeks Pembangunan Manusia	Jumlah Penduduk Miskin
1	0.424708681	0.366652809	0.2234910809	0.354992460	7.371541998
2	0.899556100	0.105630528	-0.0914161757	-0.577448007	-0.483888923
3	1.131552982	0.791453777	0.6433674229	-0.122179255	-0.404208850
4	0.137086816	0.064685857	-0.1045373114	-0.406917787	-0.350235203
5	0.183214851	-0.196336425	-0.2357486683	-0.054905866	-0.369244170
6	0.301248352	0.146575200	0.0922797240	-0.327128418	-0.040674670
7	0.436919043	0.453660236	0.5383983374	0.550554639	-0.334771950
8	0.879205497	0.376888977	0.1841276738	0.259558117	-0.269964177
9	0.268687386	0.202874123	0.0922797240	0.600618556	-0.066676573
10	0.999952412	0.100512444	-0.1176584471	0.142220810	0.196198725
11	0.162864247	-0.027439654	-0.1176584471	0.400362886	-0.039886734
12	0.700120184	0.305235802	0.1185219953	-0.061163856	0.39379821
13	0.552239131	0.719800601	0.4727926590	-0.464804192	-0.416224880
14	1.012162774	0.223346459	0.1578854024	-0.391272813	-0.473251781
15	0.165577661	-0.201454509	-0.3407177538	-0.075244333	-0.271244574
16	0.792376254	0.678855930	0.5252772017	-0.084631317	-0.359296472
17	0.137086816	0.095394360	0.0004317741	0.004545036	-0.535400268
18	0.953824377	0.776099525	0.5646406088	0.509877706	-0.375153693
19	0.994525584	0.550903831	0.3547024378	0.544296649	-0.345507584
20	-0.672867211	-0.513657629	-0.4063234323	2.454548009	-0.467834718

Showing 1 to 21 of 548 entries, 5 total columns

Modelling: Langkah ini dilakukan proses data mining dengan algoritma k-medoids. Algoritma k-medoids adalah algoritme clustering yang mirip dengan k-means. Perbedaan antara kedua algoritma tersebut adalah algortime k-medoids menggunakan objek sebagai representatif (medoid) sebagai pusat cluster dari setiap cluster, sedangkan k-means menggunakan mean sebagai pusat cluster. Kelebihan algoritma k-medoids yaitu mengatasi kelemahan algortime k-means yang sensitif pada noise dan outlier. Dalam algoritma k-means objek dengan nilai yang lebih besar diperbolehkan untuk menyimpang dari sebaran data. Keunggulan lainnya

adalah hasil dari proses *clustering* tidak bergantung pada urutan input ke *database*. Langkah-langkah algoritma *k-medoids*:

- 1) Inisialisasi *k* pusat *cluster* (jumlah *cluster*).
- 2) Pengukuran jarak *Euclidian Distance* untuk menetapkan setiap data (objek) ke *cluster* terdekat sebagai berikut:

$$d(x, y) = ||x - y||$$

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; 1, 2, 3, \dots (2.1)$$

Keterangan:

$d(x, y)$ = jarak *euclidian distance* antara pengamatan ke *i* pusat *cluster* ke *i*

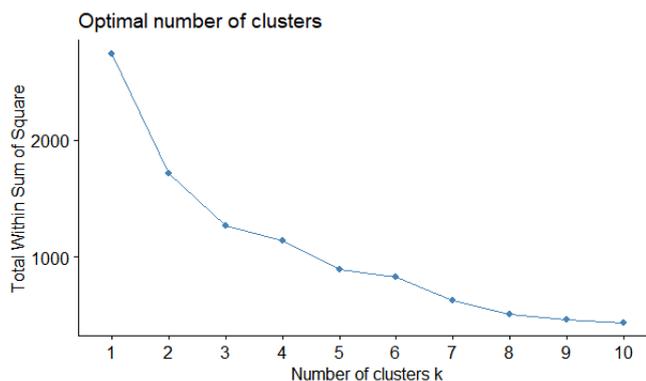
X_i = objek pada pengamatan ke *i*

Y_i = pusat kelompok ke *i*

n = banyak variabel pengamatan yang diamati

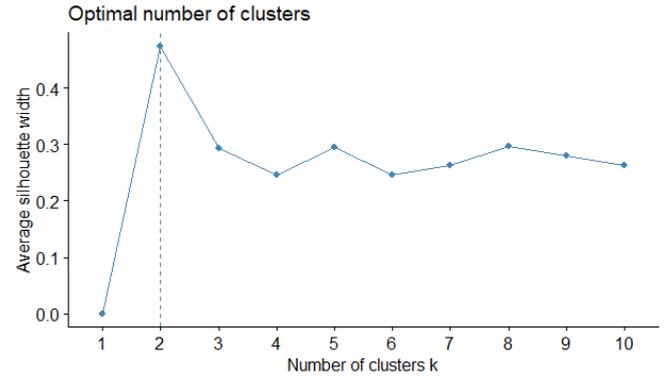
- 3) Pilih objek secara acak di setiap *cluster* sebagai kandidat untuk *medoid* baru.
- 4) Gunakan kandidat *medoid* baru untuk menghitung jarak setiap objek di setiap *cluster*.
- 5) Hitung simpangan total (*S*) dengan menghitung nilai jarak total baru – total lama. Maka total simpangan adalah $S = b - a$ Jika $S < 0$, maka tukar nilai objek dengan menentukan *medoid* baru.
- 6) Ulangi langkah 3 sampai 5 hingga tidak ada perubahan pada *medoid*, sehingga mendapatkan *cluster* dan anggota masing-masing *cluster*.

Metode *elbow*, metode *Silhouette*, metode *Factoextra* berfungsi sebagai acuan untuk pemilihan jumlah *cluster* yang optimal. Dalam pemilihan jumlah *cluster* yang optimal dengan menggunakan *Rstudio*, sehingga hasilnya dapat dilihat pada Gambar 6, Gambar 7, dan Gambar 8.



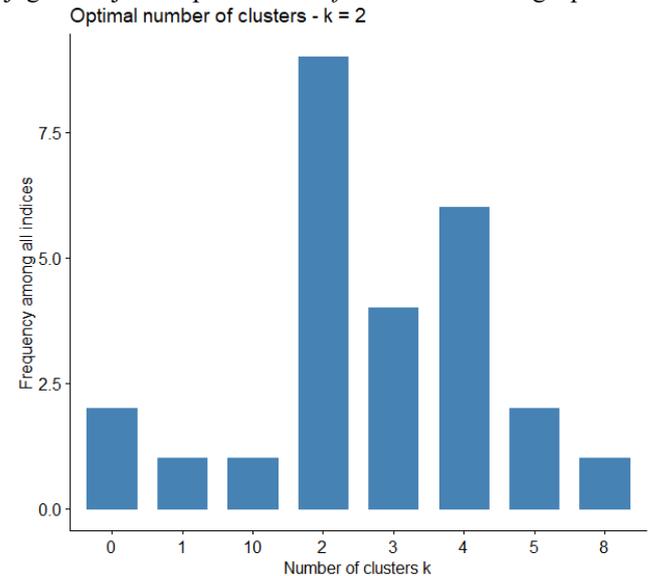
Gambar 6. Metode *Elbow* pada *Dataset* Penduduk Miskin Tahun 2020

Pada Gambar 6 digambarkan sebuah grafik dari *elbow method* yang menunjukkan nilai *k* terbaik adalah saat *k* berada di angka 2, hal ini karena nilai *k* terbaik adalah ketika terjadi penurunan di beberapa nilai *k* dan selanjutnya hasil dari nilai *k* akan stabil atau turun secara perlahan-lahan.



Gambar 7. Metode *Silhouette* pada *Dataset* Penduduk Miskin Tahun 2020

Dapat dilihat pada Gambar 7 metode *silhouette* menggunakan pendekatan nilai rata-rata untuk menduga *k* optimal dari *cluster* terbentuk, semakin tinggi nilai rata-rata maka semakin baik. Pada grafik di atas menunjukkan bahwa jumlah *cluster* optimal yang akan terbentuk adalah $k = 2$, karena $k = 2$ memiliki nilai rata-rata yang paling tinggi dibandingkan jumlah *k* lainnya. Nilai *k* optimal sebanyak 2 juga ditunjukkan pada metode *factoextra* dan fungsi *pamk*.



Gambar 8. Metode *Factoextra* pada *Dataset* Penduduk Miskin Tahun 2020

Gambar 8 menunjukkan metode *factoextra* pada *dataset* penduduk miskin tahun 2020 *cluster* optimal yaitu 2 *cluster*, karena visualisasi diatas menentukan jumlah *cluster* optimal pada *cluster* 2. Hal ini diperkuat oleh proses *clustering* dengan menggunakan fungsi *pamk()* yang menunjukkan bahwa *k* optimal berada saat $k = 2$ dapat dilihat pada Gambar 9.

kasus. Pada variabel indeks pembangunan manusia angka kasus terkecil = 4,59 kasus, angka kasus terbesar = 911,37 kasus, dengan jumlah kasus = 4790,36 kasus dan rata-rata kasus = 59,8795 kasus.

TABEL 7
RINCIAN KARAKTERISTIK *CLUSTER 2*

C1	Persentase penduduk miskin	Indeks kedalaman kemiskinan	Indeks keparahan kemiskinan	Indeks pembangunan manusia	Jumlah penduduk miskin
Min	2,02	0,07	0,04	60,48	1,46
Max	19,23	3,51	1,02	86,61	776,83
Rata-rata	8,798481	1,24222	0,278184	71,53839	51,99069
Jumlah	3765,75	1,24222	102,65	30618,43	11281,98

Cluster 2 merupakan kategori dengan tingkat persentase penduduk miskin, indeks kedalaman kemiskinan, indeks keparahan kemiskinan, indeks pembangunan manusia, dan jumlah penduduk miskin tertinggi. *Cluster 2* terdiri dari 427 kabupaten/kota. Pada variabel persentase penduduk miskin angka kasus terkecil = 2,02 kasus, angka kasus terbesar = 19,23 kasus, dengan jumlah kasus = 3765,75 kasus dan rata-rata kasus = 8,798481 kasus. Pada variabel Indeks kedalaman kemiskinan angka kasus terkecil = 0,07 kasus, angka kasus terbesar = 3,51 kasus, dengan jumlah kasus = 1,24222 kasus dan rata-rata kasus = 1,24222 kasus. Pada variabel Indeks keparahan kemiskinan angka kasus terkecil = 0,04 kasus, angka kasus terbesar = 1,02 kasus, dengan jumlah kasus = 102,65 kasus dan rata-rata kasus = 0,278184 kasus. Pada variabel Indeks pembangunan manusia angka kasus terkecil = 60,48 kasus, angka kasus terbesar = 86,61 kasus, dengan jumlah kasus = 30618,43 kasus dan rata-rata kasus = 71,53839 kasus. Pada variabel Jumlah penduduk miskin angka kasus terkecil = 1,46 kasus, angka kasus terbesar = 776,83 kasus, dengan jumlah kasus = 11281,98 kasus dan rata-rata kasus = 51,99069 kasus.

B. Pembahasan

Dalam penelitian ini untuk mengelompokkan kabupaten/kota yang memiliki kasus kemiskinan dampak covid-19 di Indonesia menggunakan algoritme *k-medoids* dan menggunakan metodologi CRISP-DM (*Cross Industry Standar Proses for Data Mining*) yang terdiri beberapa tahapan *bussines understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*.

Pada tahap *bussines understanding* dilakukan studi literatur mengenai teori-teori yang relevan dengan *data mining*, algoritme *k-medoids*, juga teknik *clustering* data kemiskinan. Permasalahan yang dianalisis peneliti pada penelitian ini yaitu terkait *clustering* data kemiskinan dampak covid-19 di Indonesia. Tahap *data understanding* merupakan tahapan untuk mengumpulkan data awal dan mempelajari data tersebut. Adapun proses pemahaman data tersebut dilakukan agar mengetahui struktur data dan penanganan terhadap data. Tahap *data preparation* merupakan tahap

dimana data yang sudah diperoleh dilakukan persiapan dan pembersihan data. Data kemiskinan yang telah dikumpulkan sebelumnya, masih berupa data yang tidak terstruktur. Oleh karena itu, diperlukan tahap *pre-processing* untuk mengubahnya menjadi data yang lebih terstruktur. Pada penelitian ini *pre-processing* terdiri dari *select data*, *data cleaning*, *data integration*, dan *format data*.

Adapun tahapan dari *preprocessing* dari *data preparation*, pertama *select data* merupakan proses pemilihan data yang akan dilakukan dalam proses *data mining (modelling)*, kemudian dilanjutkan pada tahap *data cleaning* dimana data dilakukan pengecekan apabila terdapat data yang kosong atau *missing value* maka data harus ditangani terlebih dahulu, kemudian pada tahap *data integration* melakukan penggabungan semua *dataset* yang dibutuhkan ke dalam satu tabel pada penelitian ini hanya menggunakan satu tabel, maka tahap ini tidak dilakukan, kemudian dilakukan *format data* dimana data ditransformasikan ke dalam bentuk numerik, dalam proses perubahan bentuk data dari kategorik ke numerik adalah dengan memberikan angka yang berbeda ke setiap kategorik yang berbeda untuk variabel pada *dataset*. Kemudian dilakukan standarisasi data dengan menggunakan fungsi *scale*, pada proses normalisasi menggunakan bantuan *tools Rstudio*.

Tahap *modelling* yaitu dilakukan proses *data mining* menggunakan algoritme *k-medoids* dan menggunakan *package cluster* dan *clvalid* untuk menampilkan visualisasi dari *cluster* menggunakan *package factoextra*. Kemudian dilakukan evaluasi dari hasil *modelling* atau *data mining* yang telah didapatkan yaitu untuk mengukur optimalisasi dari *cluster* yang dihasilkan pada *tools Rstudio* dengan menggunakan *package clvalid* untuk mengetahui evaluasi dari *internal measure* berupa nilai *silhouette*. Selanjutnya dilakukan *deployment* merupakan tahap pelaporan atau penyampaian hasil berupa representasi dalam bentuk yang mudah dipahami.

Adapun *cluster* optimal yang dihasilkan pada penelitian ini adalah pada *dataset* penduduk miskin tahun 2020 menghasilkan 2 *cluster* dan optimal *cluster* pada *cluster 2*. Parameter untuk mengetahui kualitas suatu *cluster* dapat dilakukan dengan menggunakan *silhouette coefficient* digunakan untuk mengukur kualitas dari *cluster*. Berdasarkan output hasil disimpulkan bahwa *cluster 2* adalah *cluster* optimal mendekati angka 1 (*Weak Structure*).

Dari *dataset* yang digunakan yaitu *dataset* penduduk miskin tahun 2020 dengan nilai *silhouette coefficient* 0,4735719. Hal tersebut menunjukkan bahwa jumlah data mempengaruhi terhadap kualitas dari sebuah *cluster*, jika jumlah data digunakan semakin banyak nilai *silhouette coefficient* yang dihasilkan semakin tinggi, dan sebaliknya jika jumlah data digunakan semakin sedikit nilai *silhouette coefficient* yang dihasilkan semakin rendah[7].

Cluster 1 merupakan kategori dengan tingkat persentase penduduk miskin, indeks kedalaman kemiskinan, indeks keparahan kemiskinan, indeks pembangunan manusia, dan jumlah penduduk miskin tertinggi. *Cluster 1* terdiri dari 121

kabupaten/kota. Pada variabel persentase penduduk miskin angka kasus terkecil = 12,92 kasus, angka kasus terbesar = 41,76 kasus, dengan jumlah kasus = 2737,14 kasus dan rata-rata kasus = 22,8095 kasus. Pada variabel indeks kedalaman kemiskinan angka kasus terkecil = 1,91 kasus, angka kasus terbesar = 13,87 kasus, dengan jumlah kasus = 566,64 kasus dan rata-rata kasus = 4,722 kasus. Pada variabel indeks keparahan kemiskinan angka kasus terkecil = 0,32 kasus, angka kasus terbesar = 6,99 kasus, dengan jumlah kasus = 174,84 kasus dan rata-rata kasus = 1,457 kasus. Pada variabel indeks pembangunan manusia angka kasus terkecil = 31,55 kasus, angka kasus terbesar = 74,46 kasus, dengan jumlah kasus = 7593,73 kasus dan rata-rata kasus = 63,2810833 kasus. Pada variabel indeks pembangunan manusia angka kasus terkecil = 4,59 kasus, angka kasus terbesar = 911,37 kasus, dengan jumlah kasus = 4790,36 kasus dan rata-rata kasus = 59,8795 kasus.

Cluster 2 merupakan kategori dengan tingkat persentase penduduk miskin, indeks kedalaman kemiskinan, indeks keparahan kemiskinan, indeks pembangunan manusia, dan jumlah penduduk miskin tertinggi. *Cluster 2* terdiri dari 427 kabupaten/kota. Pada variabel persentase penduduk miskin angka kasus terkecil = 2,02 kasus, angka kasus terbesar = 19,23 kasus, dengan jumlah kasus = 3765,75 kasus dan rata-rata kasus = 8,798481 kasus. Pada variabel Indeks kedalaman kemiskinan angka kasus terkecil = 0,07 kasus, angka kasus terbesar = 3,51 kasus, dengan jumlah kasus = 1,24222 kasus dan rata-rata kasus = 1,24222 kasus. Pada variabel Indeks keparahan kemiskinan angka kasus terkecil = 0,04 kasus, angka kasus terbesar = 1,02 kasus, dengan jumlah kasus = 102,65 kasus dan rata-rata kasus = 0,278184 kasus. Pada variabel Indeks pembangunan manusia angka kasus terkecil = 60,48 kasus, angka kasus terbesar = 86,61 kasus, dengan jumlah kasus = 30618,43 kasus dan rata-rata kasus = 71,53839 kasus. Pada variabel Jumlah penduduk miskin angka kasus terkecil = 1,46 kasus, angka kasus terbesar = 776,83 kasus, dengan jumlah kasus = 11281,98 kasus dan rata-rata kasus = 51,99069 kasus.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan hasil penelitian yang sudah dilakukan maka kesimpulan yang didapatkan adalah sebagai berikut:

1. Pada penelitian ini sudah diimplementasikan metode *clustering* menggunakan algoritme *k-medoids* untuk mengelompokkan penduduk miskin tahun 2020 dampak covid-19 di Indonesia. Dari hasil perhitungan *clustering* pada *dataset* penduduk miskin tahun 2020 menghasilkan 2 *cluster* yaitu *cluster1* dan *cluster2*. *Cluster1* terdiri 121 kabupaten/kota yang terkategori daerah dengan kasus kemiskinan dalam tingkatan tinggi. Sedangkan pada *cluster2* memiliki 427 kabupaten/kota yang terkategori sebagai daerah yang memiliki kasus kemiskinan dampak covid-19 dalam tingkatan rendah.
2. Hasil *clustering* dievaluasi dengan *silhouette coefficient* berfungsi sebagai kualitas *cluster* yang dihasilkan berdasarkan proses pengujian dari *dataset* penduduk miskin tahun 2020 menghasilkan nilai *silhouette coefficient* 0,4735719, dapat disimpulkan bahwa *cluster 2* merupakan *cluster* optimal karena mendekati angka 1 (*Weak Structure*). Hasil evaluasi dalam bentuk tabel dapat dilihat pada Tabel 4.3. Pada analisis pengelompokkan data kabupaten/kota lebih baik menggunakan *cluster 2* pada kasus kemiskinan dampak covid-19, dalam mengelompokkan kasus kemiskinan penerapan algoritme *k-medoids* mendekati ketepatan pada pengelompokkan kabupaten/kota.

B. Saran

Dari hasil penelitian yang telah dilakukan, maka saran-saran diberikan untuk mejadi referensi dalam penelitian selanjutnya, yaitu:

1. Dapat dilakukan dengan menggunakan metode *clustering* lain. Seperti algoritme *k-medoids* dan *k-means* untuk mengetahui pengklasteran mana yang lebih *robust* terhadap data yang mengandung pencilon (*outlier*). Metode *k-medoids* menggunakan *medoids* sebagai pusat klasternya dan *k-means* menggunakan *mean* sebagai pusat klasternya.
2. Diharapkan penelitian selanjutnya dapat menambah variabel yang lebih bervariasi lagi dan indeks validasi seperti *dunn* dan *connectivity*.
3. Disarankan untuk penelitian selanjutnya menggunakan *tools* yang berbeda dalam pengelompokkan wilayah yang memiliki kasus kemiskinan di Indonesia.

DAFTAR PUSTAKA

- [1] S. Internawati, "Studi Pelaksanaan Pendataan Keluarga," *Stud. Pelaks. Pendataan Kel. Miskin Dan Pemberdaya. Masy. Dalam Mengentaskan Kemiskin. Di Desa Danau Redan Kec. TelukPandan*, vol. 1, no. 1, pp. 309–323, 2013, [Online]. Available: <https://ejournal.ap.fisip-unmul.ac.id/site/?p=700>.
- [2] N. Zuhdiyaty and D. Kaluge, "Analisis Faktor - Faktor Yang Mempengaruhi Kemiskinan Di Indonesia Selama Lima Tahun Terakhir," *J. Ilm. Bisnis dan Ekon. Asia*, vol. 11, no. 2, pp. 27–31, 2018, doi: 10.32812/jibeka.v11i2.42.
- [3] B. Riyanto, "Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokkan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 562–568, 2019, doi: 10.30865/komik.v3i1.1659.
- [4] A. Y. Rofiqi, "Clustering Berita Olahraga Berbahasa Indonesia Menggunakan Metode K-Medoid Bersyarat," *J. Simantec*, vol. 6, no. 1, pp. 25–32, 2017.
- [5] R. C. Balabantaray, C. Sarma, and M. Jha, "Document Clustering using K-Means and K-Medoids," 2015, [Online]. Available: <http://arxiv.org/abs/1502.07938>.
- [6] R. D. Ramadhani and D. J. Ak, "Evaluasi K-Means dan K-Medoids pada Dataset Kecil," *Semin. Nas. Inform. dan Apl.*, no. September, pp. 20–24, 2017.
- [7] D. F. Pramesti, Lahan, M. Tanzil Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 723–732, 2017, doi:

- 10.1109/EUMC.2008.4751704.
- [8] D. E. Kurniawan and A. Fatulloh, "Clustering of Social Conditions in Batam, Indonesia Using K-Means Algorithm and Geographic Information System," *International Journal of Earth Sciences and Engineering (IJEE)*, vol. 10, no. 5, pp. 1076–1080, 2017.
- [9] P. Prasetyawan, I. Ahmad, R. I. Borman, Y. A. Pahlevi, and D. E. Kurniawan, "Classification of the Period Undergraduate Study Using Back-propagation Neural Network," 2018, pp. 1–5.
- [10] F. Mahmuda, M. A. R. Sitorus, H. Widyastuti, and D. E. Kurniawan, "Clustering Profil Pengunjung Perpustakaan Menggunakan Algoritma K-Means," *Journal of Applied Informatics and Computing*, vol. 1, no. 1, Art. no. 1, 2017, doi: 10.30871/jaic.v1i1.476.
- [11] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartana, "Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.
- [12] A. Ariyanto, D. E. Kurniawan, and A. Fatulloh, 'Rancang Bangun Aplikasi WebGIS untuk Pemetaan Kondisi Sosial Ekonomi Kota Batam', *J. Appl. Inform. Comput. JAIC*, vol. 2, no. 1, Art. no. 1, 2018, doi: 10.30871/jaic.v2i1.904.
- [13] D. Feblian and D. U. Daihani, "Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X," *J. Tek. Ind.*, vol. 6, no. 1, pp. 1–12, 2017, doi: 10.25105/jti.v6i1.1526.
- [14] B. Wira, A. E. Budianto, and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang," *RAINSTEK J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 53–68, 2019, doi: 10.21067/jtst.v1i3.3046.
- [15] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT.J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [16] S. Defiyanti, M. Jajuli, and N. Rohmawati, "K-Medoid Algorithm in Clustering Student Scholarship Applicants," *Sci. J. Informatics*, vol. 4, no. 1, pp. 27–33, 2017, doi: 10.15294/sji.v4i1.8212.
- [17] W. Budiaji and F. Leisch, "Simple k-medoids partitioning algorithm for mixed variable data," *Algorithms*, vol. 12, no. 9, pp. 1–15, 2019, doi: 10.3390/a12090177.
- [18] D. Wahyuli, H. Handrizal, I. Parlina, A. P. Windarto, D. Suhendro, and A. Wanto, "Mengelompokkan Garis Kemiskinan Menurut Provinsi Menggunakan Algoritma K-Medoids," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 452, 2019, doi: 10.30645/senaris.v1i0.51
- [19] I. Budiman, "28. 29. 1," *Data Clust. Menggunakan Metodol. Cris. Untuk Pengenalan Pola Proporsi Pelaks. Tridharma*, pp. 1–56, 2012.
- [20] R. D. Setiyawan, D. Sunaryono, and R. J. Akbar, *Rancang Bangun Aplikasi Untuk Pemetaan Tingkat Kemiskinan Masyarakat Berbasis Perangkat Bergerak*, vol. 5, no. 2, 2016.
- [21] P. Kumar and D. Sirohi, "Comparative analysis of FCM and HCM algorithm on Iris data set," *Int. J. Comput. Appl.*, vol. 5, no. 2, pp. 33–37, 2010, doi: 10.5120/888-1261.
- [23] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm for Big Data," *Phys. Procedia*, vol. 78, no. December 2015, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.
- [24] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381
- [25] D. Wahyuli, H. Handrizal, I. Parlina, A. P. Windarto, D. Suhendro, and A. Wanto, "Mengelompokkan Garis Kemiskinan Menurut Provinsi Menggunakan Algoritma K-Medoids," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 452, 2019, doi: 10.30645/senaris.v1i0.51.