

## PENERAPAN ALGORITMA *K-MEANS* DALAM PENGELOMPOKAN KESEJAHTERAAN RAKYAT DI KABUPATEN KARAWANG

**Dinda Fitriani<sup>1\*</sup>, Tesa Nur Padilah<sup>2</sup>, Betha Nurina Sari<sup>3</sup>**

<sup>1,2,3</sup>Prodi Teknik Infomatika, Universitas Singaperbangsa Karawang

<sup>1,2,3</sup> Jl. HS. Ronggo Waluyo, Telukjambe Timur, Karawang. Telp. 081295094228

\*Corresponding Author: dinda.fitriani17085@student.unsika.ac.id

### Abstrak

Untuk mempermudah pemerintah mengimplementasikan kebijakan yang terkait dengan masalah kesejahteraan rakyat, perlu diketahui daerah mana saja yang memiliki tingkat kesejahteraan rakyat rendah, sehingga pemerintah dapat memprioritaskan kesejahteraan rakyat pada wilayah tersebut. Berkaitan dengan hal tersebut, perlu dilakukan pengelompokan kesejahteraan rakyat berdasarkan wilayah yang ada. Paper ini menerapkan teknik *data mining* dengan algoritma *k-means* dalam pengelompokan kesejahteraan rakyat berdasarkan wilayah kecamatan. Hasil pengolahan data dengan bantuan *tools* Rstudio 4.0.5 diperoleh 3 *cluster*. *Cluster 1* merupakan *cluster* dengan tingkat kesejahteraan rakyat tinggi, terdiri dari 7 kecamatan. *Cluster 2* merupakan *cluster* dengan tingkat kesejahteraan rakyat sedang, terdiri dari 8 kecamatan. *Cluster 3* merupakan *cluster* dengan tingkat kesejahteraan rakyat rendah, terdiri dari 7 kecamatan. Dalam melakukan *clustering* dengan 3 *cluster* diperoleh nilai SSE sebesar 23.58788, dan tingkat akurasi sebesar 72.9%.

**Kata kunci:** Kesejahteraan Rakyat, *Data Mining*, *K-Means Clustering*

### Abstrack

*To make it easier for the government to implement policies related to people's welfare issues, it is necessary to know which areas have low levels of people's welfare, so that the government can prioritize the welfare of the people in that region. In this regard, it is necessary to classify the welfare of the people based on the existing area. This paper applies data mining techniques with the k-means algorithm in classifying people's welfare based on sub-district areas. The results of data processing with the help of tools Rstudio 4.0.5 obtained 3 clusters. Cluster 1 is a cluster with a high level of people's welfare, consisting of 7 sub-districts. Cluster 2 is a cluster with a moderate level of people's welfare, consisting of 8 sub-districts. Cluster 3 is a cluster with a low level of people's welfare, consisting of 7 sub-districts. In clustering with 3 clusters, the SSE value is 23,58788, and the accuracy rate is 72.9%.*

*Keywords: People's Welfare, Data Mining, K-Means Clustering*

### 1. Pendahuluan

Kesejahteraan merupakan tumpuan harapan dan menjadi cita-cita luhur perjuangan bangsa Indonesia, selain itu kesejahteraan merupakan hal yang menentukan suatu pembangunan di suatu daerah [1]. Kesejahteraan dapat diartikan sebagai tingkat kemampuan seseorang dalam memenuhi kebutuhan primernya seperti sandang, pangan, papan, pendidikan, dan kesehatan. Kesejahteraan bisa dinilai dari segi kesejahteraan keluarga. Kesejahteraan keluarga adalah terciptanya suatu keadaan yang harmonis dan terpenuhinya kebutuhan jasmani serta sosial bagi anggota keluarga, tanpa mengalami hambatan-hambatan yang serius di dalam lingkungan keluarga, dan dalam menghadapi masalah-masalah keluarga akan mudah untuk diatasi secara bersama oleh anggota keluarga, sehingga standar kehidupan keluarga dapat terwujud [2].

Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN) membagi kriteria keluarga sejahtera dalam tiga tahapan yakni tahapan Keluarga Pra Sejahtera (KPS), Keluarga Sejahtera I (KS 1), dan Keluarga Sejahtera II (KS II) [2]. Di Kabupaten Karawang terdapat

110.827 keluarga dengan kriteria keluarga sejahtera tahap KPS, 295.695 keluarga dengan kriteria keluarga sejahtera tahap KS I, 221.806 keluarga dengan kriteria keluarga sejahtera tahap KS II [3]. Terdapat beberapa kebijakan yang dilakukan oleh pemerintah untuk mengatasi masalah kesejahteraan rakyat di Kabupaten Karawang, mulai dari pemberian bantuan konsumtif maupun bantuan produktif, namun dalam penyalurannya sering kali mengalami masalah, sebagai contoh pada penyaluran bantuan Program Keluarga Harapan (PKH) di Kecamatan Karawang Barat Kabupaten Karawang dalam implementasinya penyaluran bantuan tersebut tidak merata, serta informasi penyaluran tidak optimal karena masih melalui informasi dari mulut ke mulut [4]. Oleh sebab itu pemerintah membutuhkan data dan informasi mengenai tingkat kesejahteraan rakyat di suatu daerah untuk memudahkan dalam mengimplementasikan kebijakannya supaya sesuai rencana dan tepat sasaran, hal tersebut dapat dilakukan pengelompokan kesejahteraan rakyat dengan menggunakan teknik *data mining*, agar diketahui daerah mana saja yang memiliki tingkat kesejahteraan rakyat rendah, sehingga pemerintah dapat memprioritaskan kesejahteraan rakyat di daerah tersebut.

*Data mining* adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. *Data mining* adalah analisis terhadap data untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya dengan cara terkini sehingga dapat dipahami dan berguna bagi pemilik data tersebut [4]. Terdapat beberapa teknik *data mining*, salah satunya yaitu *clustering*. *Clustering* digunakan untuk mengelompokkan data berdasarkan kemiripan atributnya. Metode *clustering* dapat mengelompokkan *record*, *observasi*, atau kelas yang mempunyai kesamaan objek. *Clustering* memiliki beberapa metode, salah satunya yaitu metode *clustering* non hirarki, algoritma yang sering digunakan dalam metode ini yaitu algoritma *k-means clustering*. *K-Means Clustering* merupakan algoritma yang mengelompokkan data ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain [5].

Dari permasalahan di atas, maka akan dilakukan penelitian *clustering* kesejahteraan rakyat berdasarkan kecamatan menggunakan algoritma *K-Means clustering* dengan tahap KDD (*Knowledge Discovery in Database*). *Tools* yang digunakan pada penelitian ini adalah Rstudio 4.0.5. Data yang digunakan adalah data keluarga sejahtera 2019 yang diperoleh dari Badan Pusat Statistik Kabupaten Karawang. Penelitian ini diharapkan dapat membantu pemerintah setempat untuk mengetahui kecamatan mana saja yang memiliki tingkat kesejahteraan rakyat rendah, sehingga dapat lebih memperhatikan dan berfokus pada kesejahteraan rakyat di kecamatan tersebut.

## 2. Tinjauan Pustaka

Terdapat beberapa penelitian sebelumnya yang sejalan dengan penelitian ini, diantaranya penelitian dengan judul *clustering* kabupaten/kota provinsi sulawesi tenggara berdasarkan indikator kesejahteraan rakyat [6], penelitian tersebut menggunakan algoritma *k-means* dengan bantuan *tools* WEKA 3.6. Proses *clustering* pada penelitian tersebut dimulai dari menginput *dataset* kedalam *tools* WEKA 3.6 sampai tahap terakhir yaitu menentukan *cluster* yang dilihat berdasarkan posisi objek dalam *cluster* tersebut, jika posisi objek tidak berubah maka proses *clustering* dihentikan, sehingga hasil *cluster* tersebutlah yang diambil. Hasil *clustering* yang diperoleh yaitu sebanyak 3 *cluster*, dengan kategori indikator kesejahteraan rakyat rendah, sedang, dan tinggi, yang secara berurutan beranggotakan 6 kecamatan, 3 kecamatan, dan 3 kecamatan.

Selanjutnya terdapat penelitian dengan judul penerapan *fuzzy c-means cluster* dalam pengelompokan provinsi Indonesia menurut indikator kesejahteraan rakyat [7]. Penelitian tersebut menggunakan algoritma *fuzzy c-means* dengan bantuan *tools* Rstudio, untuk melakukan pengelompokan provinsi dilakukan dengan menentukan objek *cluster* berdasarkan pada pusat *cluster* dan derajat keanggotaan pada setiap *cluster*. Hasil yang didapatkan pada penelitian ini yaitu memperoleh fungsi objektif yang konvergen pada iterasi ke-18 sebesar 130,7085 dengan 2 *cluster*. *Cluster* 1 merupakan kelompok provinsi yang kurang sejahtera atau tingkat kesejahteraan rakyat rendah dan *cluster* 2 merupakan kelompok provinsi sejahtera.

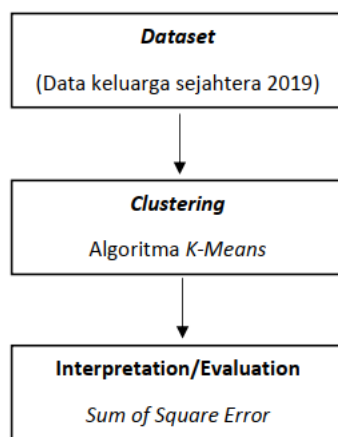
Selanjutnya terdapat penelitian dengan judul pengelompokan provinsi di Indonesia berdasarkan karakteristik kesejahteraan rakyat menggunakan metode *k-means cluster* (1855) [8]. Proses pengclusteran pada penelitian tersebut menggunakan algoritma *k-means* yang kemudian hasil dari pengclusteran dievaluasi menggunakan *Davies-Bouldin Index* (DBI)

berdasarkan nilai DBI terkecil. Proses pengolahan data pada penelitian ini dibantu dengan menggunakan Microsoft Excel, Minitab 14, dan SPSS 16. Berdasarkan pengolahan data yang dilakukan, penelitian ini memperoleh 3 *cluster* dengan nilai DBI sebesar 1,2288.

Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya, penelitian yang dilakukan saat ini yaitu penerapan algoritma *k-means* dalam pengelompokan kesejahteraan rakyat berdasarkan kecamatan di Kabupaten Karawang menggunakan metode penelitian *Knowledge Discovery in Databases* (KDD), dan proses pengolahan data dilakukan dengan *tools data mining* yaitu Rstudio.

### 3. Metodologi

Metodologi penelitian diilustrasikan pada Gambar 1, yang terdiri dari 3 bagian utama yaitu *data preparation*, *data mining*, dan *interpretation/evaluation*.



Gambar 1. Metodologi Penelitian

#### 3.1 Dataset

Data yang digunakan pada penelitian ini yaitu data klasifikasi keluarga sejahtera perkecamatan di Kabupaten Karawang pada tahun 2019. Data tersebut diperoleh dari Badan Pusat Statistik (BPS) Kabupaten Karawang. Data karakteristik keluarga sejahtera terdiri dari 3 variabel diantaranya keluarga pra-sejahtera (KPS), keluarga sejahtera I (KS I) dan keluarga sejahtera II (KS II).

#### 3.2 Clustering

Pada penelitian ini proses *clustering* dilakukan dengan bantuan *tools* Rstudio. Proses clustering dimulai dari menginput *dataset* yang diperoleh dari hasil proses *data selection*, kemudian dilakukan proses pengecekan *missing value* pada data tersebut, karena proses clustering dilakukan menggunakan *tools* Rstudio maka *dataset* perlu ditransformasi dengan cara distandarisasi (diskalakan) menggunakan fungsi *scale* pada Rstudio. Selanjutnya menentukan jumlah *cluster* optimal menggunakan metode Elbow, setelah diketahui jumlah *cluster* optimalnya kemudian dilakukan perhitungan menggunakan algoritma *K-Means*.

Berikut ini langkah-langkah yang harus dilakukan pada algoritma *K-Means Clustering* [9]:

1. Tentukan k sebagai jumlah *cluster* yang dibentuk.  
Untuk menentukan banyaknya cluster k dapat dilakukan dengan beberapa pertimbangan. Penelitian ini akan menggunakan metode Elbow untuk memilih jumlah *cluster* k yang akan digunakan untuk pengelompokan data.
2. Tentukan k *centroid* (titik pusat cluster) awal secara random.
3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *cluster*.  
Untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan *Euclidian Distance* dengan rumus sebagai berikut:

$$D(x_i, \pi_i) = \sqrt{\sum_{i=1}^n (x_i - \pi_i)^2} \quad (1)$$

Keterangan:

$(x_i, \mu_i)$  = Jarak antara *clustering*  $x$  dengan pusat *cluster*  $\mu$  pada data ke- $i$

$x_i$  = Bobot data ke- $i$  pada *cluster* yang ingin dicari jaraknya

$\mu_i$  = Bobot data ke- $i$  pada pusat *cluster*

$n$  = Jumlah data

4. Kelompokkan data berdasarkan jarak yang terdekat antara data dengan *centroid*.

5. Tentukan nilai *cluster* awal yang baru.

Untuk menentukan nilai *cluster* awal yang baru dapat dilakukan dengan cara hitung rata-rata dari *cluster* yang bersangkutan dengan rumus sebagai berikut:

$$C_k = \frac{1}{n_k} \sum_{i=1}^n d_i \quad (2)$$

Keterangan:

$n_k$  = total data dalam *cluster*  $k$

$d_i$  = total pada masing-masing *cluster*

6. Ulangi langkah tiga sampai dengan lima sampai setiap anggota *cluster* tidak ada yang berubah.

### 3.3 Interpretation/Evaluation

Setelah hasil *cluster* diperoleh dari proses *clustering*, selanjutnya dilakukan proses interpretasi *cluster*. Pada penelitian ini proses interpretasi dilakukan dengan cara menghitung nilai rata-rata setiap variabel karakteristik keluarga sejahtera pada setiap *cluster*.

Selain itu, dilakukan proses evaluasi terhadap jumlah *cluster* yang terbentuk dengan cara menghitung nilai *sum of square error*. Adapun untuk menghitung nilai SSE dapat menggunakan rumus sebagai berikut [10]:

$$SSE = \sum_{k=1}^{K=n} x_i \in S_k \|X_i - C_k\| \quad (3)$$

Keterangan:

$K$  = total *cluster*

$X_i$  = data ke- $i$

$C_k$  = awal *cluster*

$n$  = total data

## 4. Hasil dan Pembahasan

### 4.1 Dataset

Penelitian ini menggunakan data karakteristik keluarga sejahtera di Kabupaten Karawang dan disimpan dalam bentuk file Microsoft Excel. Data yang dipilih adalah data klasifikasi keluarga berdasarkan kecamatan pada tahun 2019. Variabel yang digunakan yaitu jumlah keluarga pra-sejahtera (KPS), jumlah keluarga sejahtera I (KS I), dan jumlah keluarga sejahtera II (KS II) seperti pada Tabel 1 berikut ini:

Tabel 1. Data Klasifikasi Keluarga

No	Kecamatan	KPS	KS I	KS II
1	Pangkalan	2839	4858	4004
2	Tegalwaru	2839	5566	3283
3	Ciampel	2839	7818	5716
4	Telukjambe Timur	2839	14324	10916
5	Telukjambe Barat	2839	7934	6010
6	Klari	2839	20045	17805
7	Cikampek	2839	11090	16299
8	Purwasari	2839	10114	5810
9	Tirtamulya	2839	8028	5287
10	Jatisari	2839	8968	7596

No	Kecamatan	KPS	KS I	KS II
11	Banyusari	2839	7815	5042
12	Kotabaru	2839	14507	13351
13	Cilamaya Wetan	2839	12503	9327
14	Cilamaya Kulon	2839	9340	7422
15	Lemahabang	2839	9076	4489
16	Telagasari	2839	9804	5768
17	Majalaya	2839	5554	4053
18	Karawang Timur	2694	13839	13674
19	Karawang Barat	4520	19524	11074
20	Rawamerta	3670	7390	5609
21	Tempuran	4392	9264	4828
22	Kutawaluya	4860	8069	5187
23	Rengasdengklok	5820	14549	9778
24	Jayakarta	4050	8107	5645
25	Pedes	5912	9862	7804
26	Cilebar	5266	5814	2924
27	Cibuaya	5320	8499	4604
28	Tirtajaya	6782	8134	6637
29	Batujaya	6508	10016	8027
30	Pakisjaya	2770	5284	3837

#### 4.2 Clustering

Hal pertama yang dilakukan pada tahap ini yaitu melakukan pengecekan *missing value* terhadap dataset yang sebelumnya telah diinput di Rstudio, proses pengecekan *missing value* dilakukan dengan *syntax* sebagai berikut:

```
> #pengecekan missing value
> sum(is.na(diare))
[1] 0
```

Gambar 2. Pengecekan *Missing Value*

dari Gambar 2 dapat diketahui bahwa tidak adanya *missing value* pada data yang digunakan.

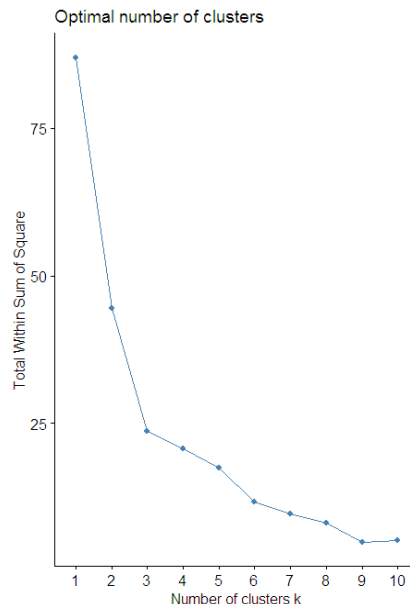
Selanjutnya *dataset* yang sudah terbebas dari *missing value* kemudian ditransformasi dengan cara distandarisasi (diskalakan) menggunakan fungsi *scale* pada Rstudio. Tabel 2 merupakan data yang sudah diskalakan/distandarisasi.

Tabel 2. Data setelah ditransformasi

No	Kecamatan	KPS	KS I	KS II
1	Pangkalan	-0.65829942	-1.314963759	-0.88527541
2	Tegalwaru	-0.65829942	-1.128709014	-1.07358557
3	Ciampel	-0.65829942	-0.536271606	-0.43813672
4	Telukjambe Timur	-0.65829942	1.1752727	0.91999483
5	Telukjambe Barat	-0.65829942	-0.505755292	-0.36135005
6	Klari	-0.65829942	2.680305742	2.71925794
7	Cikampek	-0.65829942	0.324498909	2.32592215
8	Purwasari	-0.65829942	0.067740956	-0.41358588
9	Tirtamulya	-0.65829942	-0.481026554	-0.55018257

No	Kecamatan	KPS	KS I	KS II
10	Jatisari	-0.65829942	-0.233739182	0.05288007
11	Banyusari	-0.65829942	-0.537060821	-0.61417146
12	Kotabaru	-0.65829942	1.223414816	1.55596604
13	Cilamaya Wetan	-0.65829942	0.696219183	0.50498117
14	Cilamaya Kulon	-0.65829942	-0.135876519	0.0074349
15	Lemahabang	-0.65829942	-0.205327441	-0.75860353
16	Telagasari	-0.65829942	-0.013811263	-0.4245554
17	Majalaya	-0.65829942	-1.131865874	-0.87247763
18	Karawang Timur	-0.76991039	1.047682939	1.6403269
19	Karawang Barat	0.63561802	2.543245401	0.96126113
20	Rawamerta	-0.01865314	-0.648866282	-0.46608289
21	Tempuran	0.53709248	-0.155869966	-0.6700638
22	Kutawaluya	0.89732649	-0.470240616	-0.57630049
23	Rengasdengklok	1.63626804	1.234463827	0.62277296
24	Jayakarta	0.27384455	-0.460243892	-0.45668044
25	Pedes	1.70708327	0.001446894	0.10720533
26	Cilebar	1.20983719	-1.063467239	-1.16734889
27	Cibuaya	1.25140265	-0.357119796	-0.72856793
28	Tirtajaya	2.37674905	-0.453140957	-0.19759073
29	Batujaya	2.16584282	0.041959932	0.16544828
30	Pakisjaya	-0.71141085	-1.202895226	-0.92889233

Tahap selanjutnya yaitu menentukan banyak  $k$  cluster, dengan menggunakan metode Elbow. Metode tersebut menggunakan nilai SSE (*Sum of Square Error*) untuk menentukan  $k$  optimalnya. Pada Gambar 3 merupakan grafik jumlah  $k$  cluster yang diperoleh yaitu 3 cluster.



Gambar 3. Hasil *Cluster* Optimal dengan Metode Elbow

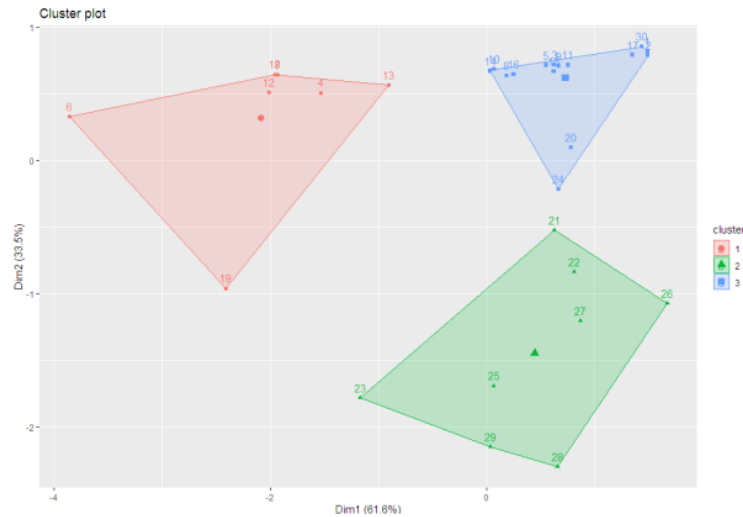
Setelah diketahui jumlah cluster yang ingin dibentuk, kemudian lakukan perhitungan data menggunakan algoritma *k-means clustering*, dengan jumlah cluster 3, sehingga

menghasilkan clustering final yang dapat dilihat pada Tabel 3. Hasil *clustering* menggunakan algoritma *k-means* dengan  $C = 3$  menghasilkan tingkat akurasi sebesar 72.9%.

Tabel 3. Hasil *Clustering* Final  $C = 3$

No	Kecamatan	Cluster
1	Pangkalan	3
2	Tegalwaru	3
3	Ciampel	3
4	Telukjambe Timur	1
5	Telukjambe Barat	3
6	Klari	1
7	Cikampek	1
8	Purwasari	3
9	Tirtamulya	3
10	Jatisari	3
11	Banyusari	3
12	Kotabaru	1
13	Cilamaya Wetan	1
14	Cilamaya Kulon	3
15	Lemahabang	3
16	Telagasari	3
17	Majalaya	3
18	Karawang Timur	1
19	Karawang Barat	1
20	Rawamerta	3
21	Tempuran	2
22	Kutawaluya	2
23	Rengasdengklok	2
24	Jayakarta	3
25	Pedes	2
26	Cilebar	2
27	Cibuaya	2
28	Tirtajaya	2
29	Batujaya	2
30	Pakisjaya	3

Plot visualisasi  $C = 3$  dapat dilihat pada Gambar 4. Pada Gambar 4, diketahui bahwa setiap *cluster* memiliki jumlah anggota yang berbeda, *cluster* 1 terdiri dari 7 anggota, *cluster* 2 terdiri dari 8 anggota dan *cluster* 3 terdiri dari 15 anggota.

Gambar 4. Plot *cluster* yang Terbentuk

#### 4.3 Interpretation/Evaluation

Untuk menginterpretasikan *cluster* yang terbentuk, maka dapat dilihat dari nilai rata-rata karakteristik tiap *cluster*, yang dapat dilihat pada Tabel 4.

Tabel 4. Data rata-rata *cluster*

<i>Cluster</i>	Rata-rata		
	KPS	KS I	KS II
1	3058	15119	13207
2	5608	9276	6224
3	2971	7710	5305

Dilihat dari Tabel 4 maka dapat diketahui bahwa *cluster* 1 merupakan *cluster* dengan tingkat kesejahteraan keluarga tinggi, *cluster* 2 merupakan *cluster* dengan tingkat kesejahteraan keluarga sedang, dan *cluster* 3 merupakan *cluster* dengan tingkat kesejahteraan keluarga rendah.

Setelah itu, melakukan evaluasi terhadap jumlah *cluster* yang terbentuk menggunakan *Sum of Square Error* (SSE). Hasilnya dapat dilihat pada Tabel 5.

Tabel 5. Perbandingan hasil SSE

Jumlah <i>Cluster</i>	Nilai SSE	Selisih SSE
2	44.53404	0
3	23.58788	20.94616
4	18.38218	5.205696
5	13.29606	5.086123
6	10.26469	3.031362
7	7.45968	2.805014
8	5.924508	1.535172
9	5.193326	0.731182
10	3.7164	1.476926

Berdasarkan Tabel 5 dapat diketahui bahwa nilai selisih SSE paling tinggi terdapat pada jumlah *cluster* sebanyak 3 *cluster*, hal tersebut membuktikan bahwa *cluster* optimal yaitu sebanyak 3 *cluster*.



#### 4.4 Pembahasan

Berdasarkan pengolahan data yang telah dilakukan menggunakan algoritma *K-Means clustering* dengan bantuan *tools* Rstudio 4.0.5 diperoleh 3 *cluster*, untuk anggota dari masing-masing *cluster* dapat dilihat pada Tabel 6.

Tabel 6. Anggota per *Cluster*

<i>Cluster</i>	Anggota (Kecamatan)
<i>Cluster 1</i>	Telukjambe Timur, Klari, Cikampek, Kotabaru, Cilamaya Wetan, Karawang Timur, Karawang Barat
<i>Cluster 2</i>	Tempuran, Kutawaluya, Rengasdengklok, Pedes, Cilebar, Cibuaya, Tirtajaya, Batujaya
<i>Cluster 3</i>	Pangkalan, Tegalwaru, Ciampel, Telukjambe Barat, Purwasari, Tirtamulya, Jatisari, Banyusari, Cilamaya Kulon, Lemahabang, Telagasari, Majalaya, Rawamerta, Jayakarta, Pakisjaya

*Cluster 1* merupakan *cluster* dengan tingkat kesejahteraan rakyat tinggi dilihat dari jumlah rata-rata keluarga pra-sejahtera (keluarga yang belum sejahtera) lebih kecil dibanding jumlah rata-rata keluarga sejahtera I dan II, dan *cluster* tersebut memiliki jumlah rata-rata keluarga sejahtera paling banyak jika dibandingkan dengan *cluster* lainnya. *Cluster 2* merupakan *cluster* dengan tingkat kesejahteraan rakyat sedang karena memiliki jumlah rata-rata keluarga sejahtera lebih banyak jika dibandingkan dengan *cluster 3*. *Cluster 3* merupakan *cluster* dengan tingkat kesejahteraan rakyat rendah karena memiliki jumlah rata-rata keluarga sejahtera paling kecil jika dibandingkan dengan *cluster* lainnya.

Dalam mengevaluasi jumlah *cluster* yang terbentuk dilakukan dengan cara menghitung nilai SSE (*Sum of Square Error*), untuk jumlah *cluster 3* memperoleh nilai SSE sebesar 23.58788 dengan tingkat akurasi 72.9%.

#### 5. Kesimpulan

Pada penelitian ini, penggunaan algoritma *k-means clustering* dalam mengelompokkan kesejahteraan rakyat berdasarkan kecamatan di Kabupaten Karawang menghasilkan 3 *cluster*, dimana kecamatan dengan tingkat kesejahteraan rakyat rendah terdapat pada *cluster 3*, kecamatan-kecamatan tersebut diantaranya Kecamatan Pangkalan, Tegalwaru, Ciampel, Telukjambe Barat, Purwasari, Tirtamulya, Jatisari, Banyusari, Cilamaya Kulon, Lemahabang, Telagasari, Majalaya, Rawamerta, Jayakarta, Pakisjaya, sehingga pemerintah dapat lebih memprioritaskan kesejahteraan rakyat di kecamatan tersebut, seperti melakukan pemberian bantuan konsumtif maupun bantuan produktif. Saran untuk penelitian selanjutnya dapat menambahkan indikator kesejahteraan rakyat lainnya agar menghasilkan tingkat akurasi yang lebih maksimal, dan penelitian selanjutnya bisa menggunakan algoritma yang berbeda seperti *k-medoids*, *fuzzy c-means*, atau dapat membandingkan hasil *clustering k-means* dengan metode *clustering* lainnya.

**DAFTAR REFERENSI**

- [1]. Soemartini, Supartini E. *Analisis K-Means Cluster Untuk Pengelompokan Kabupaten / Kota Di Jawa Barat Berdasarkan Indikator Masyarakat*. Konferensi Nasional Penelitian Matematika dan Pembelajarannya II (KNPMP II). 2017:144–154.
- [2]. Astuti, Adyatma S, Normelani E. Pemetaan Tingkat Kesejahteraan Keluarga Di Kecamatan Banjarmasin Selatan. *JPG (Jurnal Pendidikan Geografi)*. 2017;4(2):20–34.
- [3]. Cahyono B. Kabupaten Karawang Dalam Angka 2020. Karawang: Badan Pusat Statistik Kabupaten Karawang; 2020:155.
- [4]. Kurniawan A, Hakim L, Ramdani R. Evaluasi Kebijakan Program Keluarga Harapan di Kecamatan Karawang Barat Kabupaten Karawang. *Jurnal of Government and Political Studies*. 2021;4(1):40–51.
- [5]. Hakim, F. H., & Arnie, R. PERBANDINGAN PENENTUAN BUY DAN SELL PADA TRADING FOREX USDIDR DENGAN EURUSD MENGGUNAKAN METODE SUPPLY AND DEMAND DENGAN ALGORITMA K-MEANS. *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, 2019, 8(1): 23-32.
- [6]. Sy H, Rismayani, Syam A. *Data Mining Menggunakan Algoritma K-Means Pengelompokan Penyebaran Diare Di Kota Makassar*. SISITI: Seminar Ilmiah Sistem Informasi dan Teknologi Informasi. 2019:73–82.
- [7]. Tenriawaru A. Clustering Kabupaten/Kota Provinsi Sulawesi Tenggara Berdasarkan Indikator Kesejahteraan Rakyat. *semanTIK*. 2018;4(2):175–82.
- [8]. Dwitiyanti N, Selvia N, Andrari F. Penerapan Fuzzy C-Means Cluster Dalam Pengelompokan Provinsi Indonesia Menurut Indikator Kesejahteraan Rakyat. *Facto Exacta*. 2019;12(3):201–209.
- [9]. Ramdhani F, Hoyyi A, Mukid M. Pengelompokan Provinsi Di Indonesia Berdasarkan Karakteristik Kesejahteraan Rakyat Menggunakan Metode K-Means Cluster. *Jurnal Gaussian*. 2015;4(4):875–884.
- [10]. Hariyanto M, Shita RT. Penyakit DBD Menggunakan Metode Algoritma K-Means dan Metode. *SKANIKA*. 2018;1(1):117–122.
- [11]. Rahman AT, Wiranto, Rini A. Coal Trade Data Clustering Using K-Means (Case Study PT. Global Bangkit Utama). *ITSMART Jurnal Teknologi dan Informasi*. 2017;6(1):24–31.