

## Pengelompokan Tingkat Kesejahteraan Masyarakat di Sumatera Utara dengan Metode K-Means Clustering

SAGITA CHAROLINA SIHOMBING<sup>1</sup>, DINA AGNESIA SIHOMBING<sup>2</sup>

<sup>1</sup>Jurusan Fisika FMIPA, Fakultas Sains dan Teknologi, Universitas PGRI Palembang, [sagita.charolina@yahoo.com](mailto:sagita.charolina@yahoo.com)

<sup>2</sup>Program Studi Manajemen, Fakultas Ekonomi, Institut Bisnis dan Multimedia Asmi Jakarta, [agnesxya@gmail.com](mailto:agnesxya@gmail.com)

### Abstrak

Pengelompokan tingkat kesejahteraan masyarakat di Provinsi Sumatera Utara perlu dilakukan untuk memudahkan pemerintah dalam memfokuskan pembangunan pada kota/kabupaten yang tingkat kesejahteraannya masih rendah. Pada penelitian ini dilakukan pengelompokan tingkat kesejahteraan masyarakat Sumatera Utara berdasarkan beberapa variabel. Pengelompokan dilakukan dengan metode *K-means clustering*. *K-means clustering* merupakan salah satu metode clustering yang digunakan untuk mengelompok data dalam jumlah yang besar. Metode ini menghasilkan kelompok-kelompok data berdasarkan jumlah kelompok yang diinginkan. Pada penelitian ini, untuk menentukan jumlah kelompok terbaik dilakukan dengan metode Elbow. Langkah pertama dalam penelitian ini adalah membagi data kedalam kelompok-kelompok data untuk jumlah kelompok ( $k$ ) dimulai dari  $k=2$  sampai  $k=8$ . Selanjutnya, menghitung *SSE* (*Sum of Square Error*) dari klaster  $k = 2$  sampai  $k = 8$ . Setelah itu, membuat grafik Elbow dari nilai *SSE* yang dihasilkan untuk menentukan jumlah  $k$  yang paling optimal. Pengolahan data untuk mendapatkan kelompok-kelompok berdasarkan jumlah klaster ( $k$ ) dilakukan dengan Software Matlab 2013b. Kelompok data dari Software tersebut disimpan dalam Ms.excel. Sedangkan tampilan grafik Elbow yang dihasilkan dibuat dalam GUI Matlab. Dari grafik elbow yang dihasilkan tampak bahwa nilai *SSE* mengalami penurunan secara drastis pada saat  $k = 2$  ke  $k = 5$ , sedangkan dari  $k = 5$  sampai  $k = 8$ , penurunan grafik tidak signifikan. Dari sini diketahui bahwa jumlah klaster optimalnya adalah  $k = 5$ . Sehingga, dari grafik Elbow tersebut diperoleh hasil bahwa masyarakat sumatera utara secara optimal dikelompokkan dalam lima klaster. Klaster 1 hanya diisi oleh kota Medan, klaster 2 terdiri dari Kabupaten Tapanuli Utara, Kabupaten Toba, Kabupaten Simalungun, Kabupaten Dairi, Kabupaten Karo, Kabupaten Langkat, Kabupaten Humbang Hasundutan, Kabupaten Pakpak Barat, Kabupaten Samosir, Kabupaten Serdang Bedagai, Kota Padangsidimpuan, Kota Gunungsitoli, klaster 3 terdiri dari Kabupaten Deli Serdang, Kota Pematangsiantar, Kota Tebingtinggi, Kota Binjai, klaster 4 terdiri dari Kabupaten Labuhanbatu, Kabupaten Asahan, Kabupaten Batu Bara, Kabupaten Labuhanbatu Selatan, Kabupaten Labuhanbatu Utara, Kota Sibolga, Kota Tanjungbalai, dan klaster 5 terdiri dari Kabupaten Nias, Kabupaten Mandailing Natal, Kabupaten Tapanuli Selatan, Kabupaten Tapanuli Tengah, Kabupaten Nias Selatan, Kabupaten Padang Lawas Utara, Kabupaten Padang Lawas, Kabupaten Nias Utara, Kabupaten Nias Barat.

*Kata Kunci:* *K-Means Clustering*, Tingkat Kesejahteraan, Sumatera Utara, *Sum of Square Error*, Elbow

**Abstract**

*Grouping the level of community welfare in North Sumatra Province needs to be done to make it easier for the government to focus on development in cities / districts whose welfare levels are still low. In this study, the level of welfare of the people of North Sumatra was grouped based on several variables. The grouping is done using the K-means clustering method. K-means clustering is one of the clustering methods used to classify large amounts of data. This method produces groups of data based on the number of groups desired. In this study, to determine the best number of groups, the Elbow method was used. The first step in this study was to divide the data into groups of data for the number of groups ( $k$ ) starting from  $k = 2$  to  $k = 8$ . Next, calculate the SSE (Sum of Square Error) from cluster  $k = 2$  to  $k = 8$ . After that, create an Elbow graph from the resulting SSE values to determine the most optimal amount of  $k$ . Data processing to obtain groups based on the number of clusters ( $k$ ) was carried out using Matlab 2013b software. Group data from the software is stored in Ms.excel. Meanwhile, the resulting Elbow graphic display is created in the Matlab GUI. From the resulting elbow graph, it can be seen that the SSE value has decreased drastically when  $k = 2$  to  $k = 5$ , while from  $k = 5$  to  $k = 8$ , the decrease in the graph is not significant. From this we know that the optimal number of clusters is  $k = 5$ . So, from the elbow graph, the results show that the North Sumatran people are optimally grouped into five clusters. Cluster 1 is only filled by the city of Medan, cluster 2 consists of North Tapanuli Regency, Toba Regency, Simalungun Regency, Dairi Regency, Karo Regency, Langkat Regency, Humbang Hasundutan Regency, West Pakpak Regency, Samosir Regency, Serdang Bedagai Regency, Padangsidimpuan City, Kota Gunungsitoli, cluster 3 consists of Deli Serdang Regency, Pematangsiantar City, Tebingtinggi City, Binjai City, cluster 4 consists of Labuhanbatu Regency, Asahan Regency, Batu Bara Regency, South Labuhanbatu Regency, North Labuhanbatu Regency, Sibolga City, Tanjungbalai City, and cluster 5 consisting of Nias Regency, Mandailing Natal Regency, South Tapanuli Regency, Central Tapanuli Regency, South Nias Regency, North Padang Lawas Regency, Padang Lawas Regency, North Nias Regency, West Nias Regency.*

*Keywords:* K-Means Clustering, Welfare Level, North Sumatra, Sum of Square Error, Elbow

## 1. PENDAHULUAN

Tingkat kesejahteraan suatu bangsa merupakan salah satu indikator keberhasilan pembangunan suatu Negara. Peningkatan kesejahteraan yang diukur dari penurunan tingkat kemiskinan merupakan cerminan keberhasilan pembangunan yang diharapkan oleh setiap negara, termasuk Indonesia. Kemiskinan yang terjadi pada masyarakat merupakan masalah utama yang menjadi perhatian pemerintah. Sebagai suatu negara yang mempunyai cita-cita untuk mencapai bangsa yang makmur secara merata, Indonesia terus melakukan pembenahan dan Inovasi. Pembenahan tersebut dilakukan di seluruh provinsi di Indonesia, Sumatera Utara salah satunya. Sumatera Utara adalah sebuah provinsi di Indonesia yang terletak di bagian Utara Pulau Sumatera. Provinsi ini beribu kota di Medan, dengan luas wilayah 72.981,23 km<sup>2</sup> dan terdiri dari 25 kabupaten dan 8 kota. Sumatera Utara merupakan provinsi dengan jumlah penduduk terbesar keempat di Indonesia, setelah provinsi Jawa Barat, Jawa Timur, dan

Jawa Tengah (sumber: Wikipedia)[1]. Berdasarkan Rencana Pembangunan Jangka Menengah Daerah (RPJMD) Provinsi Sumatera Utara, pemerintah berusaha untuk mewujudkan kehidupan masyarakat yang maju dan sejahtera. Untuk meningkatkan kesejahteraan masyarakat Sumatera Utara, Pemerintah perlu menentukan daerah prioritas pembangunan agar tingkat kesejahteraan masyarakat dapat meningkat dan merata. Untuk mengetahui daerah prioritas tersebut, masyarakat Sumatera Utara dikelompokkan ke dalam beberapa kelompok yang memiliki karakteristik yang sama.

*Clustering* merupakan pengelompokan data berdasarkan kesamaan karakteristik masing-masing data ke dalam sejumlah kelompok (cluster) (Prasetyo, [2]). Untuk melakukan clustering (pengelompokan) ada beberapa metode yang dapat digunakan, diantaranya metode K-Means clustering. Metode ini mampu mengelompokkan data dalam jumlah yang besar dan waktu yang cepat dan efisien. Metode K-Means adalah metode klastering berbasis jarak yang membagi data ke dalam sejumlah klaster dan algoritma ini bekerja pada atribut numerik (Witten, [3]). Metode K-Means clustering merupakan metode yang sangat umum dan sederhana. Metode ini mampu mengelompokkan data dalam jumlah yang besar dan waktu yang cepat dan efisien (Asroni & Adrian, [4]). Metode K-Means Clustering digunakan dalam beberapa penelitian, diantaranya Shamrat, *et al* [5] menggunakan metode K-Means Clustering untuk menentukan jumlah permintaan beberapa jenis pekerjaan yang berbeda. Selain itu, Ali, A [6] juga menggunakan metode K-Means Clustering untuk Klasterisasi Data Rekam Medis Pasien di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo.

Untuk menentukan  $k$  terbaik, maka pada penelitian ini menggunakan metode Elbow. Metode Elbow merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah klaster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah klaster yang akan membentuk siku pada suatu titik (Madhulatha, [7]). Untuk mendapatkan perbandingannya adalah dengan menghitung nilai SSE (*Sum of Square Error*) dari masing-masing klaster. Hasil persentase yang berbeda dari setiap nilai klaster dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Jika nilai klaster pertama dengan nilai klaster kedua membentuk suatu siku dalam grafik atau nilainya mengalami penurunan paling besar, maka nilai klaster tersebut adalah yang terbaik (Bholowalia, [8]).

Oleh sebab itu, pada penelitian ini kita mengelompokkan tingkat kesejahteraan masyarakat Sumatera Utara menggunakan metode K-Means Clustering, dimana untuk menentukan kelompok yang paling optimal digunakan grafik Elbow. Tujuan dari penelitian ini adalah untuk mengelompokkan kota/kabupaten di Provinsi Sumatera Utara berdasarkan tingkat kesejahteraan masyarakat sehingga Pemerintah daerah bisa menentukan wilayah mana yang menjadi prioritas pembangunan ekonomi.

## 2. METODE PENELITIAN

Pada penelitian ini, pengelompokan data tingkat kesejahteraan masyarakat Sumatera Utara dilakukan menggunakan metode K-Means Clustering. Pengolahan data dilakukan menggunakan Software MATLAB 2013a. Data yang digunakan dalam penelitian ini adalah data tahun 2020 dan merupakan data sekunder yang diambil dari data BPS Sumatera Utara (BPS, [9]).

Pembentukan klaster tingkat kesejahteraan masyarakat Provinsi Sumatera Utara menggunakan beberapa variabel. Variabel yang digunakan terdiri dari kepadatan penduduk ( $x_1$ ) yaitu banyaknya penduduk per satuan luas daerah (jiwa per km<sup>2</sup>); jumlah pengangguran ( $x_2$ ) (dalam jiwa), yaitu jumlah mereka yang tak punya pekerjaan dan mencari pekerjaan, jumlah mereka yang tak punya pekerjaan dan mempersiapkan usaha, mereka yang tak punya pekerjaan dan tidak mencari pekerjaan, karena merasa tidak mungkin mendapatkan pekerjaan, dan mereka yang sudah punya pekerjaan, tetapi belum mulai bekerja; PDRB Perkapita atas Dasar Harga yang Berlaku ( $x_3$ ), yaitu Produk Domestik Regional Bruto (PDRB) perkapita atas dasar harga yang berlaku dihitung dengan cara Nilai PDRB dibagi jumlah penduduk dalam suatu

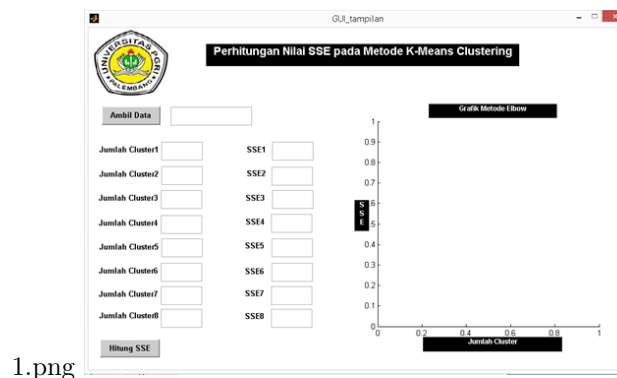
wilayah pada periode tertentu; jumlah persentase penduduk miskin ( $x_4$ ), yaitu penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah Garis Kemiskinan; Angka Partisipasi Kasar tingkat SD (APK) ( $x_5$ ), yaitu proporsi penduduk yang masih bersekolah pada suatu jenjang pendidikan tertentu terhadap jumlah penduduk kelompok umur yang sesuai dengan jenjang pendidikan tersebut; Persentase rumah tangga yang tidak memiliki fasilitas BAB ( $x_6$ ); dan Angka Harapan Hidup ( $x_7$ ), yaitu Perkiraan lama hidup rata-rata penduduk dengan asumsi tidak ada perubahan pola mortalitas menurut umur.

Pada penelitian ini, variabel yang digunakan memiliki skala yang berbeda-beda. Oleh sebab itu, kita melakukan normalisasi data dengan rumus (Hidayat, [10]):

$$x_n = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \quad (1)$$

dengan  $x_0$  merupakan nilai data aktual,  $x_{min}$  nilai minimum aktual dari seluruh data dan  $x_{max}$  nilai maksimum aktual dari seluruh data.

Normalisasi data dilakukan pada Ms. Excel. Setelah melakukan normalisasi data, langkah selanjutnya adalah membuat pengelompokan data untuk  $k = 2$  sampai  $k = 10$ . Pengelompokan dilakukan dengan aturan metode k-means clustering dan pengolahan data dilakukan dengan software Matlab 2013b. Rancangan GUI Matlab yang dibuat dalam proses pengelompokan data dan tampilan grafik Elbow. Untuk mengambil data dilakukan dengan menekan tombol Ambil Data. Selanjutnya, kita memasukkan jumlah kluster ( $k$ ) dari  $k = 1$  sampai  $k = 10$ . Hasil pengelompokan data disimpan dalam bentuk Ms.excel. Setelah itu, kita menghitung SSE (*Sum of Square Error*) dari kluster  $k = 1$  sampai  $k = 10$  dengan menekan tombol Hitung SSE. Nilai SSE untuk masing-masing  $k$  ditampilkan pada Grafik Elbow. Rancangan GUI Matlab dapat dilihat pada Gambar 1 berikut:



GAMBAR 1. Tampilan GUI Perhitungan SSE (Sum of Square Error) dan Grafik Elbow

Rancangan pada Gambar 1 digunakan untuk menampilkan grafik elbow. Grafik elbow ditampilkan berulang-ulang sampai didapat grafik yang membentuk suatu siku. Dari grafik Elbow diambil kesimpulan, jumlah kluster yang paling optimal.

### 3. HASIL DAN PEMBAHASAN

Normalisasi data menghasilkan data pada rentang 0-1, semua nilai variabel mempunyai nilai minimum 0 dan maksimum 1. Data yang sudah dinormalisasi ini, selanjutnya dikelompokkan dengan menggunakan metode K-Means Clustering. Listing program Matlab untuk pengelompokan dengan k-means clustering diberikan pada Gambar 2.

Perintah ini dibuat untuk 10 handles. Dari hasil ini diperoleh data pengelompokan tingkat kesejahteraan masyarakat Provinsi Sumatera Utara. Hasilnya terdiri dari centroid untuk masing-masing jumlah kluster yang diinput, hasil kluster, dan jarak terkecil dari centroid yang diperoleh. Data tersebut disimpan dalam suatu file Ms. Excel dengan nama hasil1.xlsx.

```

X=handles.tabel
kl=handles.kluster1
[IDX C]=kmeans(X,kl)
hasil=[X IDX]
xlswrite('hasil1.xlsx',hasil,'hasil1')
xlswrite('hasil1.xlsx',c,'kluster1')
    
```

2.png

GAMBAR 2. Listing program Matlab untuk pengelompokan dengan clustering

Selanjutnya untuk setiap nilai k yang diinput pada masing-masing jumlah kluster, dihitung SSE (*Sum of Square Error*) dengan perintah yang ditampilkan pada Gambar 3 berikut:

```

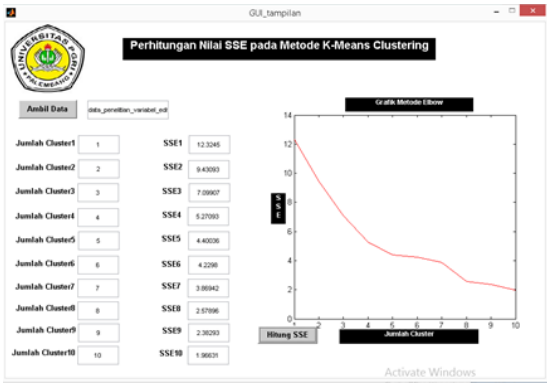
[m,n]=size(X)
for i=1:m
    for j=1:kl
        X1(i,j)=sqrt((X(i,1)-C(j,1)).^2+(X(i,2)-C(j,2)).^2+(X(i,3)-C(j,3)).^2+(X(i,4)-C(j,4)).^2+(X(i,5)-C(j,5)).^2+(X(i,6)-C(j,6)).^2+(X(i,7)-C(j,7)).^2);
    end
    Jarak1=[X1]
    xlswrite('hasil1.xlsx',Jarak1,'jarak1')

SSE1=0
for i=1:m
    SSE1=SSE1+(min(Jarak1(i,:)))^2;
end
SSE1
    
```

3.png

GAMBAR 3. Listing program Matlab untuk perhitungan SSE

Perintah ini dibuat sampai 10 SSE. Proses perhitungan dilakukan berulang-ulang sampai kita mendapat grafik elbow yang optimal. Nilai SSE (*Sum of Square Error*) terbaik yang diperoleh ditampilkan pada grafik Elbow. Tampilan grafik Elbow tampak pada Gambar 4 berikut:



4.png

GAMBAR 4. Hasil Perhitungan SSE (Sum of Square Error) dan Grafik Elbow

Dari Gambar 4, kita sudah mendapatkan grafik Elbow yang optimal dan sudah membentuk siku. Oleh sebab itu, perhitungan pada proses ini dihentikan dan hasilnya disimpan. Hasil pengelompokan data untuk jumlah kluster  $k = 2, 3, 4, 5, 6, 7, 8, 9, 10$  yang sesuai dengan Gambar 4 di atas diberikan pada Tabel 1 berikut ini:

TABEL 1. Anggota klaster untuk jumlah klaster  $k = 2, 3, 4, 5, 6, 7, 8, 9, 10$ 

Kota/Kabupaten	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
Kabupaten Nias	2	3	3	5	3	1	7	8	1
Kabupaten Mandailing Natal	2	3	1	5	2	7	4	7	4
Kabupaten Tapanuli Selatan	2	3	1	5	2	7	4	7	4
Kabupaten Tapanuli Tengah	2	3	3	5	3	1	7	8	4
Kabupaten Tapanuli Utara	2	2	2	2	1	6	2	3	2
Kabupaten Toba	2	2	2	2	1	6	2	6	3
Kabupaten Labuhanbatu	2	1	2	4	4	2	8	2	10
Kabupaten Asahan	2	1	2	4	4	2	8	2	10
Kabupaten Simalungun	2	2	2	2	1	6	2	6	3
Kabupaten Dairi	2	2	2	2	1	6	2	3	3
Kabupaten Karo	2	2	2	2	1	6	2	6	3
Kabupaten Deli Serdang	1	1	4	3	5	3	1	6	8
Kabupaten Langkat	2	2	2	2	1	6	2	3	2
Kabupaten Nias Selatan	2	3	1	5	3	1	7	8	4
Kabupaten Humbang Hasundutan	2	2	2	2	1	6	2	3	2
Kabupaten Pakpak Barat	2	2	2	2	1	6	2	3	2
Kabupaten Samosir	2	2	3	2	6	5	2	9	5
Kabupaten Serdang Bedagai	2	2	2	2	1	6	2	3	2
Kabupaten Batu Bara	2	1	2	4	4	2	8	2	10
Kabupaten Padang Lawas Utara	2	3	1	5	2	7	4	7	4
Kabupaten Padang Lawas	2	3	1	5	2	7	4	7	4
Kabupaten Labuhanbatu Selatan	2	1	2	4	4	2	8	2	10
Kabupaten Labuhanbatu Utara	2	1	2	4	4	2	8	2	10
Kabupaten Nias Utara	2	3	3	5	3	4	3	8	1
Kabupaten Nias Barat	2	3	3	5	3	4	3	8	1
Kota Sibolga	2	1	2	4	4	2	8	2	10
Kota Tanjungbalai	2	1	2	4	4	2	8	1	9
Kota Pematangsiantar	1	2	4	3	5	3	5	5	7
Kota Tebingtinggi	1	2	4	3	5	3	5	5	7
Kota Medan	1	1	4	1	5	3	6	4	6
Kota Binjai	1	2	4	3	5	3	5	5	7
Kota Padangsidimpuan	2	2	2	2	1	6	2	6	3
Kota Gunungsitoli	2	2	3	2	6	5	2	9	5

Dari Gambar 4 di atas dapat kita lihat bahwa nilai Elbow ada pada  $k = 5$  karena grafik Elbow membentuk siku pada  $k = 5$ . Sehingga pengelompokan terbaik untuk kota/kabupaten di Sumatera Utara dibuat ke dalam lima klaster. Adapun nilai centroid untuk  $k = 5$  yang sesuai dengan grafik Elbow pada Gambar 4 diberikan pada Tabel 2 di bawah ini:

TABEL 2. Nilai centroid dari setiap variabel untuk  $k = 5$ 

$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$
1.000	1.000	0.000	0.194	0.167	0.058	0.968
0,032	0,068	0,822	0,262	0,460	0,176	0,635
0,449	0,241	0,713	0,145	0,295	0,077	0,876
0,076	0,108	0,472	0,317	0,309	0,083	0,488
0.008	0.056	0.879	0.493	0.558	0.758	0.434

Hasil pengelompokan tingkat kesejahteraan masyarakat di Provinsi Sumatera Utara untuk masing-masing klaster terdiri dari kota/kabupaten yang diberikan pada Tabel 3 berikut ini:

TABEL 3. Hasil Pengelompokan Kota/Kabupaten Sumatera Utara dengan *K-Means Clustering*

Jenis Klaster	Kota/Kabupaten
Klaster 1	Kota Medan
Klaster 2	Kabupaten Tapanuli Utara, Kabupaten Toba, Kabupaten Simalungun, Kabupaten Dairi, Kabupaten Karo, Kabupaten Langkat, Kabupaten Humbang Hasundutan, Kabupaten Pakpak Barat, Kabupaten Samosir, Kabupaten Serdang Bedagai, Kota Padangsidempuan, Kota Gunungsitoli
Klaster 3	Kabupaten Deli Serdang, Kota Pematangsiantar, Kota Tebingtinggi, Kota Binjai
Klaster 4	Kabupaten Labuhanbatu, Kabupaten Asahan, Kabupaten Batu Bara, Kabupaten Labuhanbatu Selatan, Kabupaten Labuhanbatu Utara, Kota Sibolga, Kota Tanjungbalai
Klaster 5	Kabupaten Nias, Kabupaten Mandailing Natal, Kabupaten Tapanuli Selatan, Kabupaten Tapanuli Tengah, Kabupaten Nias Selatan, Kabupaten Padang Lawas Utara, Kabupaten Padang Lawas, Kabupaten Nias Utara, Kabupaten Nias Barat

Jenis klaster pada Tabel 3 merupakan pengelompokan tingkat kesejahteraan masyarakat di Provinsi Sumatera Utara. Kota/kabupaten yang berada pada klaster yang sama memiliki kemiripan karakteristik pada setiap variabel yang digunakan. Pencirian klaster untuk masing-masing variabel diberikan sebagai berikut: Klaster 1 memiliki angka Kepadatan Penduduk ( $x_1$ ) sebesar 8.603,4 jiwa per km<sup>2</sup>; jumlah pengangguran ( $x_2$ ) sebesar 94.165 jiwa; PDRB Perkapita Atas Dasar Harga Yang Berlaku ( $x_3$ ) sebesar Rp. 105.918.235,63; Jumlah Persentase Penduduk Miskin ( $x_4$ ) sebesar 8,08 %; Rata-rata Angka Partisipasi Kasar Tingkat SD ( $x_5$ ) sebesar 106,79; Persentase Rumah Tangga yang Tidak Memiliki Fasilitas BAB ( $x_6$ ) sebesar 5,37% dan Angka Harapan Hidup ( $x_7$ ) sebesar 72,98%.

Klaster 2 memiliki angka rata-rata Kepadatan Penduduk ( $x_1$ ) sebesar 317,97 jiwa per km<sup>2</sup> dengan kepadatan penduduk tertinggi pada Kota Sidempuan sebesar 1.934,7 jiwa per km<sup>2</sup> dan kepadatan penduduk terendah pada kabupaten Pakpak Barat sebesar 40,20 jiwa per km<sup>2</sup>. Rata-rata jumlah pengangguran ( $x_2$ ) pada klaster ini sebesar 6.478,58 jiwa dengan jumlah pengangguran tertinggi pada kabupaten Langkat sebesar 26.759 jiwa dan jumlah pengangguran terendah pada kabupaten Pakpak barat. Rata-rata PDRB Perkapita Atas Dasar Harga Yang Berlaku ( $x_3$ ) sebesar Rp. 36.470.261,03, pada klaster ini tidak terdapat kota/kabupaten yang PDRB Perkapita Atas Dasar Harga Yang Berlaku yang nilainya sangat tinggi ataupun sangat rendah. Rata-rata Jumlah Persentase Penduduk Miskin ( $x_4$ ) sebesar 9,55% dengan angka paling tinggi pada Kabupaten Gunungsitoli; Rata-rata Angka Partisipasi Kasar Tingkat SD ( $x_5$ ) sebesar 111,6475; Rata-rata Persentase Rumah Tangga yang Tidak Memiliki Fasilitas BAB ( $x_6$ ) sebesar 11,90% dengan nilai tertinggi pada Kota Padangsidempuan dan nilai terendah pada Kabupaten Serdang Bedagai. Untuk rata-rata angka harapan hidup ( $x_7$ ) sebesar 69,38.

Klaster 3 memiliki rata-rata Kepadatan Penduduk ( $x_1$ ) sebesar 3.885,73 jiwa per km<sup>2</sup> dengan kepadatan penduduk tertinggi pada kota tebing tinggi sebesar 5303,3 jiwa per km<sup>2</sup> dan nilai terendah pada kabupaten Deli Serdang sebesar 979,5 jiwa per km<sup>2</sup>; jumlah pengangguran ( $x_2$ ) sebesar 22.705,5 jiwa dengan jumlah pengguran terbesar pada kabupaten Deli Serdang

sebesar 60.974 jiwa dan nilai terendah pada Kota Tebingtinggi sebesar 6.831 jiwa; PDRB Perkapita Atas Dasar Harga Yang Berlaku ( $x_3$ ) sebesar Rp. 45.675.452,84; Jumlah Persentase Penduduk Miskin ( $x_4$ ) sebesar 7,03%; Rata-rata Angka Partisipasi Kasar Tingkat SD ( $x_5$ ) sebesar 108,92; Persentase Rumah Tangga yang Tidak Memiliki Fasilitas BAB ( $x_6$ ) sebesar 11,90% dan Angka Harapan Hidup ( $x_7$ ) sebesar 69,38%.

Klaster 4 memiliki rata-rata Kepadatan Penduduk ( $x_1$ ) sebesar 688,61 jiwa per km<sup>2</sup> dengan kepadatan penduduk tertinggi pada Kota Sibolga sebesar 2.121,2 jiwa per km<sup>2</sup> dan nilai terendah pada kabupaten Labuhanbatu Selatan sebesar 94,3 jiwa per km<sup>2</sup>. Rata-rata jumlah pengangguran ( $x_2$ ) sebesar 10.205,14 jiwa dengan jumlah pengangguran tertinggi pada kabupaten Asahan sebesar 22.441 dan terendah pada Kota Sibolga sebesar 3.157 jiwa; Rata-rata PDRB Perkapita Atas Dasar Harga Yang Berlaku ( $x_3$ ) sebesar Rp. 66.046.431,67 dengan nilai tertinggi pada Kabupaten Batu Bara sebesar Rp. 81.748.910,6 dan terendah pada Kabupaten Asahan sebesar Rp. 51.302.634,93; Rata-rata Jumlah Persentase Penduduk Miskin ( $x_4$ ) sebesar 10,74%; Rata-rata Angka Partisipasi Kasar tingkat SD ( $x_5$ ) sebesar 109,15; Rata-rata Persentase Rumah Tangga yang Tidak Memiliki Fasilitas BAB ( $x_6$ ) sebesar 6,76% dengan nilai tertinggi pada kabupaten Labuhanbatu Selatan sebesar 15,24% dan nilai terendah pada kabupaten Labuhanbatu sebesar 2,16%; dan rata-rata Angka Harapan Hidup ( $x_7$ ) sebesar 67,78%.

Klaster 5 memiliki rata-rata Kepadatan Penduduk ( $x_1$ ) sebesar 108,32 jiwa per km<sup>2</sup> dengan nilai tertinggi pada Kabupaten Nias Selatan sebesar 175,3 dan nilai terendah pada Kabupaten Tapanuli Selatan sebesar 46,80. Rata-rata jumlah pengangguran ( $x_2$ ) sebesar 5.363,56 jiwa dengan jumlah pengangguran terbesar pada Kabupaten Mandailing Natal sebesar 13.300 dan nilai terkecil pada Kabupaten Nias Barat sebesar 716 jiwa; PDRB Perkapita Atas Dasar Harga Yang Berlaku ( $x_3$ ) sebesar Rp. 31.658.934,26; Rata-rata Jumlah Persentase Penduduk Miskin ( $x_4$ ) sebesar 14,56%; Rata-rata Angka Partisipasi Kasar tingkat SD ( $x_5$ ) sebesar 113,27; Rata-rata Persentase Rumah Tangga yang Tidak Memiliki Fasilitas BAB ( $x_6$ ) sebesar 44,16% dan rata-rata Angka Harapan Hidup ( $x_7$ ) sebesar 67,20%.

#### 4. SIMPULAN

Grafik Elbow yang dibuat pada Gambar 4 dapat menggambarkan nilai SSE (*Sum of Square Error*) atas setiap jumlah klaster ( $k$ ) yang diinput. Tetapi, untuk membentuk suatu grafik elbow yang optimal dengan suatu siku pada salah satu  $k$  yang diinput, kita perlu melakukan perhitungan berulang-ulang. Berdasarkan grafik elbow yang dihasilkan tampak bahwa nilai SSE (*Sum of Square Error*) mengalami penurunan secara signifikan pada saat  $k = 1$  ke  $k = 5$ , sedangkan dari  $k = 5$  sampai  $k = 10$ , penurunan nilai SSE (*Sum of Square Error*) secara perlahan. Dari sini kita ketahui bahwa jumlah klaster optimalnya adalah  $k = 5$ . Sehingga, dari grafik Elbow tersebut dapat kita ketahui bahwa masyarakat Sumatera Utara secara optimal dikelompokkan dalam lima klaster. Klaster 1 hanya diisi oleh kota Medan sedangkan klaster yang lainnya diisi oleh beberapa kota/kabupaten. Dari variabel yang digunakan seperti kepadatan penduduk ( $x_1$ ); jumlah pengangguran ( $x_2$ ), PDRB Perkapita Atas Dasar Harga yang Berlaku ( $x_3$ ), jumlah persentase penduduk miskin ( $x_4$ ), Angka Partisipasi Kasar tingkat SD (APK) ( $x_5$ ), Persentase rumah tangga yang tidak memiliki fasilitas BAB ( $x_6$ ) dan Angka Harapan Hidup ( $x_7$ ) tampak bahwa kota/kabupaten yang berada pada klaster 5 memiliki tingkat kesejahteraan yang paling bawah. Oleh sebab itu, kota/kabupaten yang berada pada klaster 5 perlu mendapat perhatian dari pemerintah setempat. Adapun variabel yang perlu dipertimbangkan, yaitu Rata-rata Angka Partisipasi Kasar tingkat SD ( $x_5$ ), dan Rata-rata Persentase Rumah Tangga yang Tidak Memiliki Fasilitas BAB ( $x_6$ ). Jika dilihat dari variabel yang digunakan pada studi ini, kedua variabel ini memiliki nilai tertinggi pada klaster ini dibandingkan dengan klaster lainnya.



DAFTAR PUSTAKA

- [1] Wikipedia, 2021, *Sumatra Utara*. [https://id.wikipedia.org/wiki/Sumatra\\_Utara](https://id.wikipedia.org/wiki/Sumatra_Utara)
- [2] Prasetyo, E., 2012, *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*, Andi, Yogyakarta.
- [3] Witten, et al., 2012, *Data Mining Practical Machine Learning Tools and Technique, 2nd Edition*, Morgan Kaufmann, San Francisco.
- [4] Asroni & Adrian, R., 2015, Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik. *Jurnal Ilmiah Semesta Teknik*, Volume 18.
- [5] Shamrat, F. J. M., Tasnim, Z., Mahmud, I., Jahan, M. N., & Nobel, N. I., 2020, Application of K-means clustering algorithm to determine the density of demand of different kinds of jobs. *International Journal of Scientific & Technology Research*, 9(02), 2550-2557.
- [6] Ali, A., 2019, Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 19(1), 186-195.
- [7] Madhulatha, T.S., 2012, An Overview on Clustering Methods. *IOSR Journal of Engineering*, II(4), pp.719-725
- [8] Bholowalia, P., & Kumar, A., 2014, EBK-Means: A Clustering Techniques based on Elbow Method and K-Means in WSN. *International Journal of Computer Application (0975-8887)*, IX(105), 17-24.
- [9] Badan Pusat Statistik (BPS), 2020, *Provinsi Sumatera Utara dalam Angka 2020*. <https://sumut.bps.go.id/publication/2020/04/27/317f98717fcca50650c40477/provinsi-sumatera-utara-dalam-angka-2020.html>
- [10] Hidayat, R., Suprpto., 2012., Meminimalisasi nilai error peramalan dengan algoritma extreme learning machine. *Jurnal Optimasi Sistem Industri*, 11(1), 187-192.

