

Pengelompokan Status Ekonomi Keluarga Desa Tanjungsari menggunakan Metode *K-Means* *Clustering*

FIQRI FAKHRUL GUNAWAN, FAJRI RAKHMAT UMBARA, FATAN KASYIDI

Fakultas Sains dan Informatika Universitas Jenderal Achmad Yani, Cimahi, Indonesia

Email: figrifakhrulq18@if.unjani.ac.id

Received 26 Agustus 2022 | *Revised* 2 November 2022 | *Accepted* 24 Desember 2022

ABSTRAK

Pada tahun 2020 hingga sampai saat ini tahun 2022 telah terjadi fenomena pandemi yang menyebabkan penurunan ekonomi yang cukup signifikan sehingga perubahan ekonomi masyarakat berubah, pengelompokan data harus dilakukan dengan teknik yang baik karena akan berpengaruh terhadap hasil akhir pengelompokan. Sehingga penelitian ini dilakukan untuk menginterpretasi kelompok yang terbentuk dari implementasi k-means clustering menggunakan 3 teknik similarity yaitu Euclidean, Manhattan dan Minkowski Distances yang memiliki nilai kemurnian tinggi berdasarkan nilai dari silhouette coefficient serta nilai cluster yang ditentukan pada penelitian ini menggunakan Teknik elbow method. Penelitian ini menghasilkan 5 cluster yang dihasilkan dari elbow method. Dengan menghasilkan nilai silhouette coefficient dari euclidean 0.059, manhattan 0.0946, dan minkowski 0.059.

Kata kunci: *status ekonomi, data mining, K-mean Clustering, silhouette coefficient, Euclidean Distance, Manhattan Distance, Minkowski Distance*

ABSTRACT

In 2020 until now in 2022 there has been a pandemic phenomenon which has caused a significant economic decline so that changes in the community's economy have changed, data grouping must be done with good technique because it will affect the final result of the grouping. So this research was conducted to interpret the groups formed from the implementation of k-means clustering using 3 similarity techniques namely Euclidean, Manhattan and Minkowski Distances which have a high purity value based on the value of the silhouette coefficient and the cluster values determined in this study using the elbow method technique. This study produced 5 clusters resulting from the elbow method. By producing a silhouette coefficient value of euclidean 0.059, manhattan 0.0946, and minkowski 0.059.

Keywords: *economic status, data mining, K-mean Clustering, silhouette coefficient, Euclidean Distance, Manhattan Distance, Minkowski Distance*

1. PENDAHULUAN

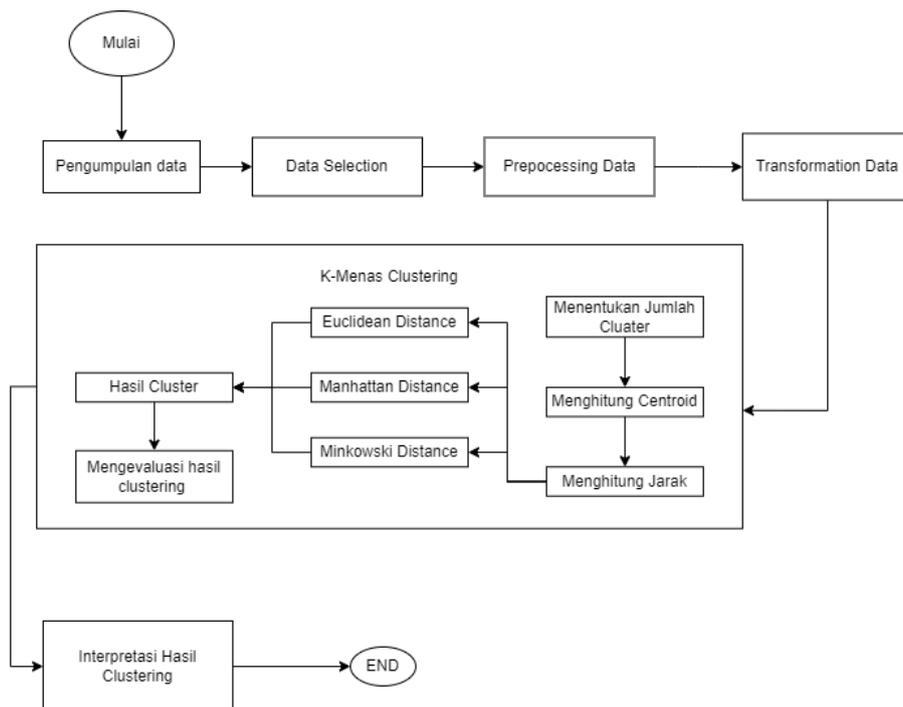
Pada tahun 2020 hingga sampai saat ini tahun 2022 telah terjadi fenomena pandemi yang menyebabkan penurunan ekonomi yang cukup signifikan sehingga perubahan ekonomi masyarakat berubah. Bapak Muhadjir Effendy yang saat ini menjabat sebagai Menko PMK mengatakan, banyak masyarakat yang status ekonominya disebut rendah tetapi tidak menerima bantuan dari pemerintah atau distribusi tidak merata. Padahal mereka harusnya mendapatkan bantuan tersebut (**Novrizaldi, 2021**).

Pada suatu penelitian terdahulu (**Edison et al., 2021**) penelitian ini menghasilkan klusterisasi status ekonomi penduduk dengan menghasilkan 3 kluster yaitu tinggi, sedang, dan rendah. Dengan menggunakan atribut penghasilan, pengeluaran, dan jumlah tanggungan. pada penelitian tersebut menentukan klusternya 3 secara langsung, tetapi pada konsep clustering menggunakan k means clustering belum tentu cluster tersebut merupakan cluster terbaik. Pada penelitian lain yang berjudul "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering" menghasilkan dari *Euclidean distance*, *Manhattan distance*, dan *Minkowski distance* yang di evaluasi menggunakan silhouette coefficient menghasilkan *Euclidean Distance* menghasilkan nilai 0,51, *Manhattan distance* menghasilkan 0,516, dan *minkowski distance* menghasilkan 0,52. Dengan menggunakan data dengan kasus disparitas kebutuhan guru (**Hidayati et al., 2021**). Pada penelitian yang lain mengenai perbandingan performa antara *Euclidean distance*, *Manhattan distance*, dan *Minkowski distance* menghasilkan *Minkowski Distance* dengan nilai 16.3418, *Euclidean Distance* 18.3532, dan *Manhattan Distance* 29.4712. dengan kasus pengelompokan tingkat kejahatan pada kejahatan tinggi, sedang, dan rendah di wilayah Bontang (**Haviluddin et al., 2020**). Maka dari itu pada penelitian ini akan menggunakan 3 teknik perhitungan jaraknya yaitu *Euclidean*, *Manhattan* and *Minkowski Distances* terhadap kasus status ekonomi keluarga.

Dalam beberapa penelitian yang menggunakan K-means pada penjelasan diatas, ditarik kesimpulan bahwa penggunaan metode k means clustering dalam penentuan nilai k atau nilai cluster yang dipilih tidak bisa sembarangan karena nilai cluster tersebut belum tentu merupakan cluster terbaik dari data yang ada maka terdapat metode elbow yang berfungsi untuk menentukan nilai kluster terbaik dalam metode k-means clustering. Pada k-means clustering juga terdapat beberapa Teknik similarity yang merupakan tehnik penentuan *distance* pada k-means yang hasil dari pengelompokannya dapat di evaluasi menggunakan silhouette coefficient sehingga nilai cluster terbaiknya dapat di lihat dari score yang dihasilkan. Berdasarkan pembahasan di paragraf sebelumnya, maka dalam penelitian ini melakukan pengelompokan status ekonomi keluarga di desa tanjungsari menggunakan Kmeans Clustering. Dalam pengerjaannya menggunakan 3 teknik *similarity* yaitu *Euclidean*, *Manhattan* and *Minkowski Distances* terhadap status ekonomi keluarga. Serta dalam penentuan nilai clusternya pada penelitian ini menggunakan elbow method untuk penentuan nilai clusternya agar penentuan nilai cluster tersebut merupakan jumlah cluster terbaik. Untuk mengevaluasi hasil cluster yang dilakukan maka pada penelitian ini menggunakan metode *silhouette coefficient* yang menghasilkan nilai silhouette dari setiap cluster yang dilakukan.

2. METODE PENELITIAN

Selanjutnya langkah-langkah yang akan dilakukan pada penelitian ini dapat dilihat pada Gambar 1 dibawah ini.



Gambar 1. Alur Metode Penelitian

2.1. Pengumpulan Data

Pada penelitian ini menggunakan data yang diambil dari kantor desa Tanjungsari. Data tersebut merupakan data yang diambil langsung dari warga melalui petugas desa atau melalui program SDGs desa dimana SDGs Desa merupakan Langkah awal untuk pembangunan berkelanjutan yang merupakan program paling utama penggunaan Dana Desa Tahun 2021. Data yang diambil merupakan data masing-masing rumah tangga di Desa Tanjungsari yang terdiri dari 1200 record data.

2.2. Data Selection

Data Selection adalah proses untuk memilih atau menseleksi data dari sumber data yang sesuai dengan tujuan dari penelitian atau hanya mengambil data data yang dibutuhkan (Mardi, 2017). Data yang dipilih dalam penelitian ini memiliki atribut data sebagai berikut :

Tabel 1. Tabel Deskripsi Data

No. 1	Nama Atribut 2	Keterangan 3
1.	Nama Kepala Keluarga	Merupakan merupakan nama dari kepala keluarga yang menjadi narasumber dalam data yang digunakan.
2.	RT	Merupakan alamat Rukun Tetangga(RT) dari rumah tangga tersebut.
3.	RW	Merupakan alamat Rukun Warga(RW) dari rumah tangga tersebut

No. 1	Nama Atribut 2	Keterangan 3
4.	Tempat Tinggal yang Ditempati	Merupakan status tempat tinggal rumah tangga tersebut yang di tempati
5.	Status lahan tempat tinggal yang ditempati	Merupakan status lahan dari tempat tinggal yang di tempati oleh rumah tangga
6.	Luas Lantai	Merupakan luas dari rumah yang di tempati yang memiliki satuan meter persegi (m^2)
7.	Luas Lahan	Merupakan luas dari Lahan yang di tempati yang memiliki satuan meter persegi (m^2)
8.	Jenis Lantai tempat tinggal terluas	Merupakan jenis lantai dari tempat tinggal yang di tempati
9.	Dinding sebagian besar rumah	Merupakan jenis dinding rumah yang ditempati
10.	Jendela	Merupakan status jendela yang ada di rumah yang ditempati
11.	Atap	Merupakan jenis atap yang digunakan pada tempat tinggal yang ditempati
12.	Penerangan	Merupakan jenis penerangan rumah yang digunakan pada tempat tinggal
13.	Energi Untuk Memasak	Merupakan jenis energi untuk memasak
14.	Fasilitas MCK	Merupakan kondisi MCK yang ada di rumah tangga tersebut
15.	Sumber Air Mandi	Merupakan sumber air bersih yang digunakan untuk mandi
16.	Tempat Buang air besar	Merupakan status tempat buang air besar yang ada di rumah tangga tersebut
17.	Sumber air minum terbanyak	Merupakan sumber air minum terbanyak yang ada di rumah tangga
18.	Penghasilan	Merupakan penghasilan dari kepala rumah tangga dari keluarga tersebut

2.3. Preprocessing Data

Tahapan *pre-processing* adalah proses mengubah data mentah menjadi format yang dapat dipahami. Tahapan ini merupakan langkah yang penting dalam melakukan data mining, karena data mining tidak dapat dilakukan apabila data yang ada masih merupakan data yang mentah. Ada beberapa proses yang dilakukan pada tahap *pre-processing* diantaranya:

a. Data Cleaning

Data Cleaning atau pembersihan data merupakan tahap awal dalam melakukan proses data mining (**Mardi, 2017**). Dalam proses pembersihan data ini bisa dilakukan untuk menghapus data yang kosong atau missing value.

2.4. Transformasi Data

Data *Transformation* yaitu tahapan yang digunakan untuk mengubah bentuk data yang belum dilakukan apa-apa menjadi data numerik untuk algoritma *K-Means Clustering*. Perubahan data yang dilakukan menyesuaikan dengan penelitian sebelumnya yang menggubakan data yang hampir sama (**Taruna & Setiadi, 2020**). penentuan angka pada transformasi yang dilakukan berdasarkan tingkat harga beli dari suatu atribut contohnya Diding Sebagian besar rumah data "Semen/beton/kayu berkualitas tinggi" di beri nilai 1 dan "Kayu berkualitas rendah/bamboo" diberi nilai 1 karena harga beli dari semen/beton/kayu berkualitas tinggi lebih tinggi dibandingkan dengan Kayu berkualitas rendah/bamboo. Berikut perubahan data yang dilakukan :

Tabel 2. Tabel Tranformasi

NO 1	Inisialisasi 2	Atribut 3	Keterangan 4	Nilai 5
1	Q1	Luas Lantai yang Ditempati	(tidak Berubah)	
2	Q2	Luas Lahan yang Ditempati	(tidak Berubah)	
3	Q3	Tempat Tinggal yang ditempati	Milik Sendiri	6
			Kontrak/Sewa	5
			Bebas Sewa	4
			Dipinjam	3
			Dinas	2
			Lainnya	1
4	Q4	Status Tempat Tinggal Yang Ditempati	Milik Sendiri	4
			Milik Orang Lain	3
			Tanah Negara	2
			Lainnya	1
5	Q5	Jenis Tempat Tinggal Terluas	Marmmer/granit	8
			Keramik	7
			Parket/vinil/permadani	6
			Ubin/tegel/teraso	5

NO 1	Inisialisasi 2	Atribut 3	Keterangan 4	Nilai 5
			Kayu/papan kualitas tinggi	4
			Semen/bata merah	3
			Bambu	2
			Kayu/papan kualitas rendah	1
6	Q6	Dinding Sebagian besar rumah	Semen/beton/kayu berkualitas tinggi	2
			Kayu berkualitas rendah/bamboo	1
7	Q7	Jendela	Ada, Berfungsi	3
			Ada, Tidak berfungsi	2
			Tidak ada	1
8	Q8	Atap	Genteng	3
			Kayu/Jerami	2
			Lainnya	1
9	Q9	Penerangan Rumah	Listrik PLN	5
			Listrik non PLN	4
			Lampu minyak/lilin	3
			Sumber penerangan lainnya	2
			Tidak ada	1
10	Q10	Energi untuk memasak	Gas kota/LPG/biogas	3
			Minyak tanah/batu bara	2
			Kayu bakar	1
11	Q11	Fasilitas MCK	Sendiri	4
			Berkelompok/tetangga	3
			MCK Umum	2
			Tidak ada	1
12	Q12	Sumber air mandi terbanyak	Ledeng/perpipaan berbayar/air isi ulang/kemasan	5
			Perpipaan	4
			Mata air/ sumur	3
			Sungai, danau, embung	2
			Tadah air hujan	1
13	Q13	Fasilitas buang air besar	Jamban Sendiri	3

NO 1	Inisialisasi 2	Atribut 3	Keterangan 4	Nilai 5
			Jamban bersama/tetangga	2
			Jamban umum	1
14	Q14	Sumber air minum terbanyak	Ledeng/perpipaan berbayar/air isi ulang/kemasan	4
			Mata air/perpipaan/sumur	3
			Sungai, danau, embung	2
			Tadah air hujan	1
15	Q15	Penghasilan	(Tidak Berubah)	

Tahapan berikutnya melakukan normalisasi data pada setiap data dengan menggunakan teknik data *scaling* yaitu *min-max normalization*.

Tabel 3. Normalisasi *min-max Normalization*

No 1	Nama Kepala Keluarga 2	Q1 3	Q2 4	Q3 5	Q4 6	...	Q15 17
1	Imas Sumartin	0.190	0.087	1	1	...	0.064
2	Komar Suheherman	0.190	0.087	1	1	...	0.064
3	Shinta Ramyani Suherman	0.133	0.057	1	1	...	0.081
4	Rahmat Saepudin	0.133	0.067	1	1	...	0.064
5	Ade Kausar Arifin	0.133	0.047	1	1	...	0.064
6	Iwan Darmawan	0.133	0.047	1	1	...	0.039
7	Kunaefi Rahmat	0.077	0.047	1	1	...	0.047
8	Adang Bahtiar	0.133	0.047	0.8	0.6666 67	...	0.097
9	Deni Suryadi	0.133	0.082	1	1	...	0.064
10	Edi	0.073	0.027	1	1	...	0.632
...
1197	Endang S	0.062	0.037	1	0.334	...	0.037

2.5. Proses K-Means Clustering

Setelah melakukan tahapan *Pre-Processing* terhadap data yang akan digunakan maka Langkah berikutnya melakukan penerapan metode data mining, algoritma data mining yang digunakan untuk melakukan pengelompokan ini dengan menggunakan algoritma *K-Means Clustering* kepada data yang sudah melewati tahap *Pre-Processing* dan *transformation*.

2.4.1. Penentuan Jumlah Cluster

Pada tahap penentuan jumlah *Cluster* yang akan di pilih ini menggunakan metode Elbow, Metode Elbow yaitu teknik untuk menentukan jumlah kelompok atau *cluster* dimana *cluster* tersebut merupakan cluster terbaik dilihat dari nilai *Sum Of Squere Error* (SSE) (**Aditya et al., 2020**). Menggunakan method *Elbow Method* ini akan menghasilkan jumlah klaster terbaik sebelum melakukan klasterisasi. Jadi pada penelitian ini menggunakan *elbow method* untuk menghasilkan jumlah k terbaik terhadap pengelompokan status ekonomi keluarga. Metode elbow method menggunakan rumus *Sum of Squere Error* :

$$SEE = \sum_{k=1}^k \sum_{x_i \in C_k} \|x_i - c_k\|_2^2 \quad (1)$$

Keterangan :

N = banyaknya cluster,

x_i = data ke-i,

c_k = nilai rata rata cluster.

2.4.2. Menghitung *Centroid*

Pada tahap menghitung *centroid* di tentukan menggunakan rata rata dari data yang digunakan sesuai dengan hasil dari *elbow method* yang dihasilkan.

2.4.3. Menghitung Jarak Antara Objek

Pada penelitian ini karena melihat pengaruh 3 teknik *Similarity* terhadap status ekonomi keluarga maka pada tahap perhitungan jarak menggunakan 3 teknik *similarity* yaitu Euclidean, Manhattan, minkowski.

1. Euclidean *Distance*

Penghitungan jarak dengan Teknik *similarity* pertama menggunakan *Euclidean Distance* yaitu satu cara dalam menghitung jarak yang berfungsi untuk menghitung jarak 2 (dua) buah data dalam ruang Euclidean (**Nishom, 2019**). *Euclidean Distance* dapat diterapkan dengan rumus sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Keterangan :

d = Jarak antara titik x dan dan titik y,

x = data utama klaster,

y = data pada atribut,

i = seluruh data,

n = jumlah data,

x_i = data pada kelaster utama ke-i ,

y_i = data pada setiap data ke-i.

2. Manhattan *Distance*

Perhitungan jarak Teknik *similarity* yang kedua yaiitu menggunakan *Manhattan Distance* yang dihasilkan atas dasar penjumlahan perbedaan jarak antara dua *class* lalu nilai yang didapatkan dari *Manhattan Distance* bernilai bulat atau mutlak (**Pribadi et al., 2022**). Dengan rumus sebagai berikut :

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (3)$$

Keterangan :

- d = Jarak antara titik x dan dan titik y ,
- x = data utama klaster,
- y = data pada atribut,
- i = seluruh data,
- n = jumlah data,
- x_i = data pada kelaster utama ke- i ,
- y_i = data pada setiap data ke- i .

3. Minkowski *Distance*

Penghitungan jarak yang ketiga menggunakan Teknik similarity *Minkowski Distance* yang merupakan generalisasi dari jarak Euclidean dan jarak Manhattan (**Nishom, 2019**), dengan rumus sebagai berikut :

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p} \quad (4)$$

Keterangan :

- d = Jarak antara titik x dan dan titik y ,
- x = data utama klaster,
- y = data pada atribut,
- i = seluruh data,
- n = jumlah data,
- x_i = data pada kelaster utama ke- i ,
- y_i = data pada setiap data ke- i .
- p = power.

2.4.4. Hasil Klaster

Setelah melakukan perhitungan jarak data dengan centroid telah diketahui hasilnya maka selanjutnya kelompokan data berdasarkan klaster dan menghasilkan sebuah *cluster* dari setiap Teknik *similarity*, Pengelompokan data tersebut dapat dibedakan berdasarkan jarak terpendek dengan dihitung menggunakan 3 teknik *similarity* Euclidean, Manhattan, dan minkowski.

2.4.5 Evaluasi

Untuk merancang evaluasi dari hasil cluster ini menggunakan Metode *Silhoutte Coeffisien*. Metode ini berfungsi untuk menguji keakurasian nilai dari cluster yang dihasilkan (**Zhihong & Jie, 2019**). Teknik *Silhoutte Coeffisien* bekerja melakukan pengecekan cluster yang menyatukan metode *cohesion* dan *Separation* (**Tambunan et al., 2020**). Tahapannya sebagai berikut :

1. Menghitung jarak rata-rata antara objek dengan semua objek lain yang berada di dalam satu cluster.
2. Menghitung jarak rata-rata antara objek tersebut dengan objek lain di cluster lain, kemudian diambil nilai minimumnya.
3. Langkah terakhir hitung menggunakan rumus *Silhouette Coefficient*.

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (5)$$

Keterangan :

$a(i)$ = jarak rata-rata objek i ke semua objek lain dalam *cluster* yang sama.

$b(i)$ = jarak rata-rata objek i dari semua objek lain di *cluster* lain.

2.5 Interpretasi Hasil *Clustering*

Tahapan interpretasi hasil dari clustering ini mendeskripsikan hasil dari pengelompokan yang ada, bagaimana data yang dihasilkan, menentukan Pola-pola dalam suatu kelompok akan memiliki kemiripan ciri/sifat terhadap pola-pola dalam kelompok yang lainnya.

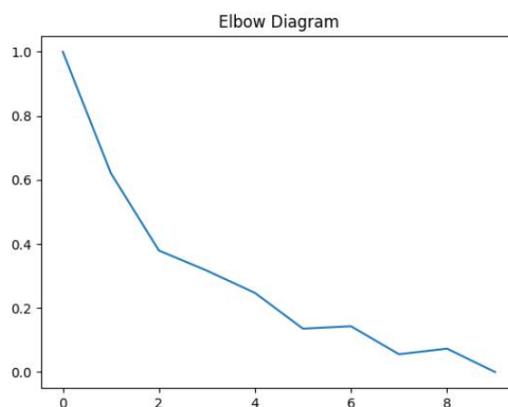
3. HASIL DAN PEMBAHASAN

Penjelasan dari penelitian ini menghasilkan berapa jumlah cluster yang dihasilkan dari elbow method, *score silhouette coefficient* mana yang terbesar antara *Euclidean*, *Manhattan*, dan *minkowski*. Hasil interpretasi data yang dihasilkan.

3.1 Penentuan Jumlah Cluster

Penentuan jumlah cluster yang dihasilkan menggunakan elbow method yaitu sebanyak 5 cluster, bisa dilihat pada Gambar 2 di bawah ini :

Hasil Dari Elbow



Gambar 2. Visualisasi hasil eksperimen

Elbow method adalah grafik yang melengkung ke suatu titik atau paling menurun nilainya sehingga jumlah nilai cluster benar dan dapat dilihat pada sudut terkecil dari grafik yang dihasilkan. **(Muningsih & Kiswati, 2018)**. Dapat dilihat dari Gambar 2 diatas dapat dilihat grafik perpotongan yang memiliki sudut terkecil yaitu di angka klaster 5 karena gambar grafik selanjutnya mulai stabil.

3.2. Hasil Evaluasi dari *Euclidean*, *Manhattan*, *Minkowski*

Perhitungan yang telah dilakukan menggunakan metode *K-Means Clustering* selanjutnya melakukan pengujian terhadap performace dengan memeriksa kualitas hasil pengelompokan

untuk setiap perhitungan jarak. Pengujian dihitung dengan persamaan *Silhouette coefficient*. Penentuan nilai k itu ditentukan oleh *elbow method*, setelah melakukan perhitungan menggunakan sistem yang telah dibuat, hasil perhitungan elbow method ada di angka 5 yang merupakan perpotongan menuju nilai yang stabil. Jadi pada saat melakukan perhitungan klasterisasi nilai k yang di tentukan atau jumlah cluster yang ditentukan yaitu 5. Dengan hasil dari *silhouette score* dapat dilihat pada Tabel 4

Tabel 4. Score Silhoutte

<i>Distance</i> 1	Jumlah Kelompok 2	<i>Score Silhoutte</i> 3
Euclidean <i>Distance</i>	4	0.0962
Manhattan Distancce	4	0.1413
Minkowski <i>Distance</i>	4	0.0962
Euclidean <i>Distance</i>	5	0.0590
Manhattan Distancce	5	0.0946
Minkowski <i>Distance</i>	5	0.0590
Euclidean <i>Distance</i>	6	0.2410
Manhattan Distancce	6	0.1590
Minkowski <i>Distance</i>	6	0.2410

Berdasarkan hasil perhitungan *Silhouette Coeficiente* yang tertera pada Tabel 4 dapat dilihat nilai perhitungan score silhouete pada *Manhattan Distance* yaitu 0.0946 nilai ini merupakan nilai paling besar dibandingkan dengan *Euclidean Distance* dan *Manhattan Distance*.

Ada beberapa faktor yang mempengaruhi nilai yang di hasilkan dari *Silhouette Coeficiente* seperti tahap preprocessing, nilai centroid yang di tentukan secara acak, dan jumlah cluster yang di tentukan. Pada hasil eksperimen yang telah di lakukan factor yang menjadikan nilai silhouette yaitu jumlah cluster yang di tentukan. Apabila jumlah clusternya semakin banyak maka nilai solhoutte yang dihasilkan juga semakin kecil atau jika jumlah cluster semakin kecil maka nilai *silhouttenya* semakin besar.

Adapun faktor yang menyebabkan nilai *silhoutte coefficient* dari Manhattan *distance* yaitu faktor data yang digunakan. Data yang digunakan memiliki nilai dimensi yang tinggi atau jaraknya jauh dan banyak sehingga sulit untuk ditafsirkan dan divisualisasikan dengan menggunakan *meuclidean distance*, sedangkan menggunakan Manhattan *distance* faktor tersebut dapat di kendalikan karena mengambil jumlah nilai absolut dari perbedaan antara koordinat.

3.3 Interpretasi Data Cluster

Adapun Hasil Dari Interpretasi yang telah dilakukan dapat di lihat pada Tabel 5 berikut :

Tabel 5. Interpretasi Data (1)

	Q1 1	Q2 2	Q3 3	Q4 4	Q5 5	Q6 6	Q7 7	Q8 8	Q9 9	Q10 10
1	Luas Lantai > 60 m	Luas Lahan > 60	Milik Sendiri	Milik Sendiri	Keramik	Beton	Ada, Berfungsi	Genteng	Listrik PLN	Gas
C1	71%	93%	93%	92%	52%	69%	98%	98%	100%	100%

C2	7%	23%	76%	79%	16%	85%	97%	97%	100%	100%
C3	34%	44%	89%	89%	11%	82%	97%	98%	100%	100%
C4	97%	90%	95%	97%	14%	90%	98%	98%	100%	100%
C5	30%	48%	66%	69%	14%	66%	97%	98%	100%	100%

Tabel 4. Interpretasi Data (2)

	Q11 11	Q12 12	Q13 13	Q14 14	Q15 15
	Sendiri	Mata air / sumur	Jamban Sendiri	Isi Ulang	Penghasilan > 24 Jt/thn
C1	97%	72%	96%	94%	66%
C2	99%	0%	22%	89%	74%
C3	99%	1%	23%	91%	26%
C4	99%	2%	21%	91%	76%
C5	99%	4%	21%	68%	54%

Dari interpretasi diatas, maka disimpulkan bahwa cluster 1 memiliki makna Status ekonominya yang paling tinggi, cluster 2 memiliki makna status ekonominya ke 3 tertinggi, cluster 3 memiliki makna yang status ekonominya ke 4 tertinggi, cluster 4 memiliki makna status ekonominya tertinggi ke 2, dan cluster 5 memiliki makna status ekonominya yang terendah yang ada di Desa Tanjungsari.

4. KESIMPULAN

Hasil dari pengelompokan yang dilakukan terhadap penelitian ini yaitu berjumlah 5 *cluster* atau kelompok, yang dihasilkan dari *Elbow Method* yang memiliki kemurnian tinggi berada di angka 5. Setelah melakukan pengelompokan menggunakan 3 teknik similarity, hasil cluster tersebut dievaluasi menggunakan silhouette coefficient dengan menghasilkan *score silhouette* dari *Euclidean Distance* yaitu sebesar 0.059, *score silhouette* dari *Manhattan Distance* yaitu sebesar 0.094, dan *score silhouette* dari *Minkowski Distance* yaitu sebesar 0.059. Maka *Score Silhoutte* terbesar dari 3 teknik similarity tersebut menggunakan 5 cluster yaitu menggunakan *Manhattan Distance*. Maka hasil pengelompokan status ekonomi keluarga di desa tanjungsari menggunakan metode k-means clustering yang terbaik menggunakan Teknik similarity *Manhattan Distance*. Alasan mengapa Teknik Manhattan yang lebih tinggi karena berada di faktor data yang digunakan memiliki nilai dimensi yang tinggi atau jaraknya jauh dan banyak sehingga sulit untuk ditafsirkan dan divisualisasikan dengan menggunakan euclidean *distance*, sedangkan menggunakan Manhattan *distance* faktor tersebut dapat di kendalikan karena mengambil jumlah nilai absolut dari perbedaan antara koordinat.

UCAPAN TERIMA KASIH

Pertama-tama penulis ingin menyampaikan terima kasih kepada Kantor Desa Tanjungsari yang telah menyediakan bahan data kepada penulis untuk digunakan pada penelitian ini. Terimakasih juga kepada berbagai pihak telah memberikan pula masukan, bantuan, *support system* serta masukan dan tekanan yang baik secara moril maupun materil sampai penelitian

ini terselesaikan. Semoga penelitian yang penulis lakukan dapat memberikan kontribusi bagi perkembangan keilmuan khususnya pada bidang teknologi.

DAFTAR RUJUKAN

- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, 4(1), 51. <https://doi.org/10.30865/mib.v4i1.1784>
- Edison, E., Yupianti, Y., & Elfianty, L. (2021). Application of the K-Means Method in Determining the Clustering of the Economic Status of the Villagers of Gunung Megang. *Jurnal Komputer, Informasi Dan ...*, 50–58. <https://penerbitadm.com/index.php/KOMITEK/article/view/230%0Ahttps://penerbitadm.com/index.php/KOMITEK/article/download/230/427>
- Haviluddin, Iqbal, M., Putra, G. M., Puspitasari, N., Setyadi, H. J., Dwiyanto, F. A., Wibawa, A. P., & Alfred, R. (2020). A Performance Comparison of Euclidean, Manhattan and Minkowski Distances in K-Means Clustering. *2020 6th International Conference on Science in Information Technology: Embracing Industry 4.0: Towards Innovation in Disaster Management, ICSITech 2020*, 184–188. <https://doi.org/10.1109/ICSITech49800.2020.9392053>
- Hidayati, R., Zubair, A., Pratama, A. H., & Indana, L. (2021). Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering. *Techno.Com*, 20(2), 186–197. <https://doi.org/10.33633/tc.v20i2.4556>
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Muningsih, E., & Kiswati, S. (2018). Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan. *Joutica*, 3(1), 117. <https://doi.org/10.30736/jti.v3i1.196>
- Nishom, M. (2019). Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(1), 20–24. <https://doi.org/10.30591/jpit.v4i1.1253>
- Novrizaldi. (2021). *No Title*. Kementerian Koordinator Bidang Pembangunan Manusia Dan Kebudayaan Republik Indonesia. <https://www.kemenkopmk.go.id/temukan-bansos->

- tidak-merata-di-kantong-kemiskinan-menko-pmk-data-lapangan-harus-disempurnakan
Pribadi, W. W., Yunus, A., Wiguna, A. S., Malang, K., Timur, J., Distance, E., & Distance, M. (2022). *PERBANDINGAN METODE K-MEANS EUCLIDEAN DISTANCE DAN MANHATTAN DISTANCE PADA PENENTUAN ZONASI COVID-19 DI. 6(2)*.
- Tambunan, H. B., Barus, D. H., Hartono, J., Alam, A. S., Nugraha, D. A., & Usman, H. H. H. (2020). Electrical peak load clustering analysis using K-means algorithm and silhouette coefficient. *Proceeding - 2nd International Conference on Technology and Policy in Electric Power and Energy, ICT-PEP 2020*, 258–262. <https://doi.org/10.1109/ICT-PEP50916.2020.9249773>
- Taruna, R. M., & Setiadi, T. A. P. (2020). Penentuan Rumus Percepatan Tanah Akibat Gempabumi Di Kota Mataram Menggunakan Metode Euclidean *Distance*. *JST (Jurnal Sains Dan Teknologi)*, 9(1), 20–29. <https://doi.org/10.23887/jst-undiksha.v9i1.23613>
- Zhihong, L., & Jie, Z. (2019). *Online Knowledge Community Governance Based on Blockchain Token Incentives*. https://doi.org/10.1007/978-981-15-1209-4_5