

PERBANDINGAN METODE K-MEANS EUCLIDEAN DISTANCE DAN MANHATTAN DISTANCE PADA PENENTUAN ZONASI COVID-19 DI KABUPATEN MALANG

Waskito Wahyu Pribadi, Amak Yunus, Anggri Sartika Wiguna

Program Studi Teknik Informatika S1, Fakultas Sains dan Teknologi

Universitas PGRI Kanjuruhan Malang

Jl. S. Supriadi No. 48, Bandungrejosari, Kec. Sukun, Kota Malang, Jawa Timur, Indonesia

wahyu.pribadi6789@gmail.com

ABSTRAK

Coronavirus (Corona Virus Disease) atau yang biasa disebut Covid-19 ini adalah sebuah penyakit yang sangat berbahaya. Banyaknya jumlah kasus Covid-19 yang tercatat di Kabupaten Malang membuat pemerintah dituntut untuk membagi Zonasi wilayah pada setiap Kecamatan di Kabupaten Malang agar dapat membuat kebijakan atau peraturan yang dapat ditaati oleh zona-zona tertentu. K-Means adalah sebuah algoritma untuk mengelompokkan data berdasarkan titik pusat kluster (*Centeroid*) yang paling dekat dengan data tersebut [1]. K-Means bertujuan untuk mengelompokkan data dengan memaksimalkan kesamaan data dalam kluster dan meminimalkan kesamaan data antar kluster [2]. Kualitas hasil *Clustering* yang lebih baik dari *Euclidean Distance* dan *Manhattan Distance* pada metode K-Means diharapkan dapat menentukan *Clustering* pada Zonasi Covid-19 di Kabupaten Malang. Metode K-Means dengan *Euclidean Distance* memperoleh nilai *Silhouette Coefficient* dan Standart Deviasi yang lebih baik dari metode K-Means *Manhattan Distance* dengan perbandingan nilai *Silhouette Coefficient* $0,71 > 0,64$ dan nilai Standart Deviasi $0,46466002 < 0,4961977$.

Kata kunci: *Corona Virus Disease, Euclidean Distance, Manhattan Distance*

1. PENDAHULUAN

Penyakit yang sangat berbahaya yaitu Covid-19 atau yang biasa dikenal sebagai Virus Corona (*Coronavirus Disease*). Pada bulan Desember 2019 kota Wuhan, China merupakan kota pertama yang terjangkit oleh Covid-19 [3]. 7.736 kasus Covid-19 terkonfirmasi pada tanggal 30 Januari 2020 di China [3]. Disisi lain, pertanggal 29 Maret 2020 sebanyak 1.528 kasus terkonfirmasi di Indonesia. Banyaknya jumlah kasus Covid-19 yang tercatat di Kabupaten Malang pemerintah dituntut untuk membagi Zonasi wilayah pada setiap Kecamatan di Kabupaten Malang agar dapat membuat kebijakan atau peraturan yang harus ditaati oleh zona-zona tertentu.

Pembagian Zonasi Covid-19 di Kabupaten Malang dapat berpengaruh terhadap pengambilan kebijakan yang dilakukan oleh pemerintah. Adapun penentuan Zonasi Covid-19 dapat berpengaruh terhadap kegiatan masyarakat di Kabupaten Malang khususnya di sektor perdagangan, pariwisata, pendidikan, pusat perbelanjaan dan lain-lain.

Pengelompokan wilayah di kabupaten malang tentunya tidak terlepas dari cara pemilihan metode pengelompokan (*Clustering*). Ada berbagai macam metode dalam pengelompokan, salah satu metode non hierarki yang dapat mempartisi atau mengumpulkan data ke dalam beberapa kluster disebut dengan metode K-Means [1].

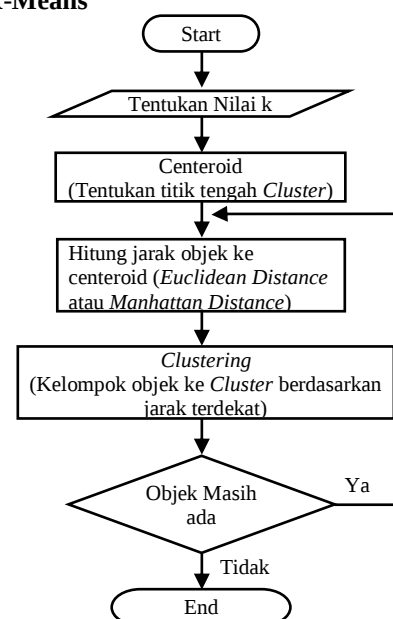
K-Means adalah sebuah algoritma untuk mengelompokkan data berdasarkan titik pusat kluster (*Centeroid*) yang paling dekat dengan data tersebut [1]. K-Means bertujuan untuk mengelompokkan data dengan memaksimalkan kesamaan data dalam kluster dan meminimalkan kesamaan data antar kluster [2].

Kualitas hasil yang lebih baik dari *Euclidean Distance* dan *Manhattan Distance* pada metode K-Means diharapkan dapat menentukan *Clustering* pada Zonasi Covid-19 di Kabupaten Malang.

Adanya tujuan dalam penelitian ini adalah membandingkan perhitungan *K-Means Euclidean Distance* dan *Manhattan Distance* menggunakan Standar Deviasi dan *Silhouette Coefficient* yang diharapkan dapat menentukan *Clustering* pada Zonasi Covid-19 di Kabupaten Malang.

2. TINJAUAN PUSTAKA

2.1. K-Means



Gambar 1. Flowchart Diagram K-Means [4]

K-Means adalah sebuah algoritma untuk mengelompokkan data berdasarkan titik pusat kluster (*Centeroid*) yang paling dekat dengan data tersebut [1]. K-Means bertujuan untuk mengelompokkan data dengan memaksimalkan kesamaan data dalam kluster dan meminimalkan kesamaan data antar kluster [2].

1. Menentukan nilai k sebagai jumlah *Cluster* yang ingin di bentuk.
2. Menentukan nilai acak atau random untuk pusat *Cluster* awal centeroid sebanyak k , untuk menghitung jarak setiap data input terhadap masing-masing centeroid dengan menggunakan rumus jarak (*Euclidean Distance* atau *Manhattan Distance*).
3. Hitung jarak setiap data ke masing-masing centeroid menggunakan persamaan (*Euclidean Distance* atau *Manhattan Distance*).
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan centeroidnya.
5. Tentukan posisi centeroid baru (k).
6. Ulangi kembali langkah ke 3 jika posisi centeroid baru dengan centeroid lama tidak sama atau data berpindah kelompok [5].

2.2. Euclidean Distance

Euclidean Distance merupakan salah satu metode perhitungan jarak yang digunakan untuk mengukur jarak dari 2 (dua) buah titik dalam euclidean space (meliputi bidang Euclidean dua dimensi, tiga dimensi atau bahkan lebih) [6]. rumus *Euclidean Distance* untuk mengukur tingkat kemiripan data adalah sebagai berikut:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad [6]$$

Keterangan:

d = jarak antara x dan y

x = data pusat kluster

y = data pada atribut

i = setiap data

n = jumlah data

x_i = data pada pusat kluster ke i

y_i = data pada setiap data ke i

Euclidean Distance mempunyai kelebihan dan kekurangan dibandingkan dengan metode perhitungan lain. Kelebihan dalam *Euclidean Distance* adalah sebagai berikut:

1. Perhitungan jarak menggunakan *Euclidean Distance* lebih umum digunakan dalam metode *K-Means*.
2. *Euclidean Distance* mempunyai hasil yang lebih optimal dibandingkan dengan perhitungan yang lainnya.
3. Referensi dari *Euclidean Distance* lebih banyak karena secara umum perhitungan *Euclidean Distance* digunakan dalam perhitungan jarak pada *K-Means Clustering*.

Kekurangan dari *Euclidean Distance* adalah sebagai berikut:

1. Perhitungan manual menggunakan *Euclidean Distance* cenderung lebih lama karena jumlah iterasi *Euclidean Distance* lebih banyak dibandingkan dengan *Manhattan Distance* [7].

2.3. Manhattan Distance

Manhattan Distance digunakan untuk menghitung perbedaan absolut (mutlak) antara koordinat sepasang objek [6]. Rumus yang digunakan adalah sebagai berikut:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad [6]$$

Keterangan:

d = jarak antara x dan y

x = data pusat kluster

y = data pada atribut

i = setiap data

n = jumlah data

x_i = data pada pusat kluster ke i

y_i = data pada setiap data ke i

Manhattan Distance mempunyai kelebihan dan kekurangan dalam perhitungan metode *K-Means*. Kelebihan dari *Manhattan Distance* adalah sebagai berikut:

1. Rata-rata kesalahan kuadrat diantara nilai actual dan nilai peramalan (squared error) menggunakan Bank dataset dan diuji menggunakan tool WEKA menunjukkan bahwa *Manhattan Distance* lebih baik dibandingkan *Euclidean Distance* [8].
2. *Manhattan Distance* cenderung lebih cepat dibandingkan *Euclidean Distance* pada perhitungan manual karena jumlah iterasi pada *Manhattan Distance* cenderung lebih sedikit dalam perhitungan manual [7].

Kekurangan dari *Manhattan Distance* adalah sebagai berikut:

1. *Manhattan Distance* mempunyai tingkat akurasi yang lebih lemah dibandingkan dengan *Euclidean Distance* dengan presentase tertinggi adalah *Euclidean Distance* mencapai 100% sedangkan *Manhattan Distance* hanya mencapai 80% [9].

2.4. Silhouette Coefficient

Silhouette Coefficient adalah penggabungan dari metode *cohesion* dan *separation* yang merupakan metode evaluasi untuk *Cluster*. Jarak antara data dapat dihitung menggunakan rumus *Euclidean Distance* dan *Manhattan Distance*. untuk mendapatkan informasi tentang kualitas hasil *Clustering* pada proses *Clustering* dapat dihitung menggunakan *silhouette* dari masing-masing *Cluster* bahkan *Cluster* dari hasil kerja algoritma *Clustering* [10]. Pada metode ini terdapat beberapa tahapan yaitu sebagai berikut:

1. Menghitung rata-rata jarak dari suatu dokumen misalkan I dengan semua dokumen lain yang berada dalam satu *Cluster*.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)$$

Dengan j adalah dokumen lain dalam satu Cluter A dan $d(i, j)$ adalah jarak antara dokumen i dan j .

2. Hitung rata-rata jarak dari dokumen I tersebut dengan semua dokumen di *Cluster* lain dan diambil nilai terkecilnya.

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j)$$

Dengan $d(i, C)$ adalah jarak rata-rata dokumen i dengan semua objek pada *Cluster* lain C dimana $A \neq C$.

$$b(i) = \min_{C \neq A} d(i, C)$$

3. Nilai *Silhouette Coefficient* nya adalah:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

2.5. Standar Deviasi

Standar Deviasi atau simpangan baku merupakan salah satu cara mengukur variasi sekelompok data kuantitatif. Besar nilai standar deviasi menunjukkan tingkat varians dari sekumpulan data. Semakin besar nilai standar deviasi, berarti semakin bervariasi atau heterogen angka-angka pada data kuantitatif tersebut. Standar deviasi (Standar Deviasi) digunakan lambang SD. Rumus umum standar deviasi adalah sebagai berikut:

$$SD = \sqrt{\frac{\sum x^2}{N}}$$

Keterangan:

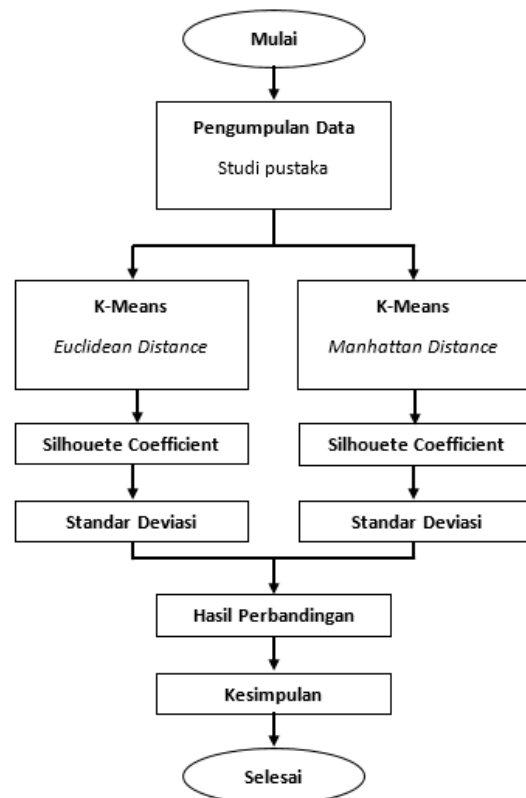
SD : Standar Deviasi
 x : $X - M_x$ = deviasi
 $\sum x^2$: Jumlah deviasi yang telah dikuadratkan
 N : banyaknya angka/nilai x [11].

3. METODE PENELITIAN

3.1. Rancangan Penelitian

1. Penelitian ini menggunakan data yang diperoleh dari hasil studi pustaka pada website resmi Satgas Covid-19. Data yang diperoleh dari hasil studi pustaka yaitu data Kecamatan dan data pasien Covid-19 pada bulan Oktober 2020 hingga bulan Juli 2021.
2. Setelah data diperoleh, maka penulis menggunakan metode K-Means dengan perhitungan jarak *Euclidean Distance* dan *Manhattan Distance*.
3. *Silhouette Coefficient* dan Standar Deviasi digunakan untuk memperoleh kualitas hasil *clustering* dari masing-masing model metode K-Means.
4. Hasil dari *Silhouette Coefficient* dan standar deviasi dapat digunakan untuk menentukan mana

yang lebih bagus dari kedua model metode K-Means.



Gambar 2. Rancangan penelitian

3.2. Metode Pengumpulan Data

Secara umum proses atau prosedur dapat digunakan untuk memecahkan suatu masalah disebut dengan Metode. Studi pustaka merupakan metode yang digunakan penulis untuk mengambil informasi sebagai dasar untuk menyelesaikan penelitian ini.

3.3. Studi Pustaka

Studi Pustaka adalah metode pengumpulan data yang diarahkan kepada pencarian data dan informasi melalui dokumen, baik dokumen tertulis, gambar, maupun dokumen elektronik yang dapat mendukung dalam proses penulisan.

Studi pustaka yang dikaji adalah jumlah penduduk dari 33 Kecamatan yang berada di Kabupaten Malang sebanyak 2.935.138 orang. Data kedua adalah jumlah pasien positif Covid-19 yang berada di Kabupaten Malang sebanyak 30.460 pasien dari bulan Oktober 2020 sampai dengan bulan Juli 2021.

3.4. Populasi dan Sampel

Data yang digunakan oleh penulis dalam melakukan penelitian ini adalah seluruh data Kecamatan dan data pasien Covid-19 di Kabupaten Malang. Sampling jenuh merupakan bagian dari *nonprobability* sampling yang digunakan penulis

sebagai Teknik Sampling. Sampling jenuh digunakan karena data populasi digunakan sepenuhnya sebagai data sampel penelitian [12]. Data yang digunakan adalah 2.935.138 orang untuk jumlah penduduk dengan total 33 Kecamatan di Kabupaten Malang dan 30.460 pasien untuk jumlah pasien Covid-19 di Kabupaten Malang.

3.5. Analisis Data

Penelitian ini menggunakan teknik analisis data yang diperoleh dari hasil studi pustaka atau mengumpulkan data dan informasi melalui website resmi dari satgas covid-19. Selanjutnya data akan diolah dan dideskripsikan agar dapat memperoleh hasil yang diinginkan oleh penulis. Berikut adalah tahapan-tahapan yang dilakukan oleh penulis:

A. Metode K-means dengan model *Euclidean Distance*

1. Metode K-Means yang menggunakan model perhitungan jarak *Euclidean Distance* digunakan untuk mengelompokkan data pada beberapa kluster yang sudah ditentukan. Data yang digunakan dalam rumus *Euclidean Distance* ini menggunakan data Kecamatan di Kabupaten Malang sebanyak 33 Kecamatan dan data pasien Covid-19 di Kabupaten Malang sebanyak 30.460 pasien.
2. Pada rumus *Euclidean Distance*:

$$d = \sqrt{(\text{centeroid 1} - \text{data 1})^2 + (\text{centeroid 2} + \text{data 2})^2}$$

Data centeroid 1 dan centeroid 2 diambil dari data pasien yang dipilih secara random. Kemudian dihitung pada masing-masing data pasien pada bulan pertama dan bulan kedua hingga keseluruhan data sudah dihitung.

3. Menghitung *Silhouette Coefficient* dan Standart deviasi untuk mendapatkan informasi tentang kualitas hasil pada metode K-means dengan model *Euclidean Distance*.

B. Metode K-Means dengan model *Manhattan Distance*

1. Metode K-Means dengan model *Manhattan Distance* digunakan untuk mengelompokkan data pada beberapa kluster yang sudah ditentukan. Data yang digunakan dalam rumus *Manhattan Distance* ini menggunakan data Kecamatan di Kabupaten Malang sebanyak 33 Kecamatan dan data pasien Covid-19 di Kabupaten Malang sebanyak 30.460 pasien.
2. Pada rumus *Manhattan Distance*:

$$d = |\text{Centeroid 1} - \text{data 1}| + |\text{Centeroid 2} - \text{data 2}|$$

Data centeroid 1 dan centeroid 2 diambil dari data pasien yang dipilih secara random. Kemudian dihitung pada masing-masing data pasien pada bulan pertama dan bulan kedua hingga keseluruhan data sudah dihitung.

3. Menghitung *Silhouette Coefficient* dan Standart deviasi untuk mendapatkan informasi tentang kualitas hasil pada metode K-means dengan model *Manhattan Distance*.

4. HASIL DAN PEMBAHASAN

4.1. Data Kecamatan dan Data Pasien Covid-19

Perbandingan metode K-means model *Euclidean Distance* dan *Manhattan Distance* pada pasien Covid-19 di Kabupaten Malang menggunakan data Kecamatan sebanyak 33 Kecamatan dan data pasien Covid-19 pada bulan Oktober 2020 - Juli 2021 sebanyak 30.460 pasien. Berikut adalah data dari Kecamatan dan Pasien Covid-19 di Kabupaten Malang:

1. Data Kecamatan

Tabel 1. Data Kecamatan di Kabupaten Malang

No	Kecamatan	No	Kecamatan
1	Donomulyo	18	Pakis
2	Pagak	19	Pakisaji
3	Bantur	20	Ngajum
4	Sumbermanjing Wetan	21	Wagir
5	Dampit	22	Dau
6	Ampelgading	23	Karangploso
7	Poncokusumo	24	Singosari
8	Wajak	25	Lawang
9	Turen	26	Pujon
10	Gondanglegi	27	Ngantang
11	Kalipare	28	Kasembon
12	Sumberpucung	29	Gedangan
13	Kepanjen	30	Tirtoyudo
14	Bululawang	31	Kromengan
15	Tajinan	32	Wonosari
16	Tumpang	33	Pagelaran
17	Jabung		

Tabel di atas merupakan daftar kecamatan yang berada di Kabupaten Malang. Data Kecamatan tersebut diambil dari Website Resmi Satgas Covid-19 di Kabupaten Malang.

2. Data Pasien Covid-19

Tabel 2 merupakan daftar pasien yang terjangkit Covid-19 pada bulan Oktober 2020 hingga bulan Juli 2021. Data tersebut di ambil dari website resmi Satgas Covid-19 di Kabupaten Malang. Pengambilan data tersebut di rangkum pada setiap akhir bulan di tahun 2020 hingga 2021.

Tabel 2. Data pasien Covid-19

No	Kecamatan	Bulan 10	Bulan 11	Bulan 12	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6	Bulan 7
1	Donomulyo	18	18	18	23	27	40	62	82	123	247
2	Pagak	6	6	9	10	18	26	33	35	42	150
3	Bantur	9	11	14	16	24	37	47	50	55	149
4	Sumbermanjing Wetan	10	12	12	18	26	29	32	36	43	184
5	Dampit	22	28	40	45	61	74	86	92	108	290
6	Ampelgading	5	5	5	6	9	15	15	15	21	43
7	Poncokusumo	10	13	21	33	36	42	49	55	59	145
8	Wajak	16	20	20	35	43	43	47	48	54	174
9	Turen	19	23	36	74	93	102	110	125	135	415
10	Gondanglegi	22	29	32	44	60	73	79	84	93	174
11	Kalipare	12	13	15	16	22	39	42	48	53	164
12	Sumberpucung	8	13	21	35	50	66	78	81	43	208
13	Kepanjen	70	81	95	116	161	189	228	263	309	790
14	Bululawang	39	41	44	63	69	77	86	89	93	212
15	Tajinan	17	21	25	29	32	37	43	49	55	109
16	Tumpang	21	25	32	38	45	49	54	64	72	151
17	Jabung	22	24	24	35	41	45	48	56	60	160
18	Pakis	102	115	158	197	238	241	264	298	318	720
19	Pakisaji	53	59	84	101	113	124	134	153	166	433
20	Ngajum	9	9	12	18	24	37	50	55	61	184
21	Wagir	38	46	50	65	70	77	84	93	106	304
22	Dau	46	54	67	86	107	124	134	141	145	465
23	Karangploso	60	74	84	127	154	175	197	213	220	480
24	Singosari	204	243	289	328	380	404	439	478	524	1220
25	Lawang	113	121	143	201	240	261	290	319	355	774
26	Pujon	16	22	34	49	53	59	60	61	61	131
27	Ngantang	13	14	17	27	28	33	33	34	49	271
28	Kasembon	1	3	6	8	8	11	11	11	11	263
29	Gedangan	6	8	9	10	11	14	15	16	18	61
30	Tirtoyudo	4	4	7	12	14	16	22	28	37	240
31	Kromengan	13	14	18	22	30	49	54	59	63	165
32	Wonosari	3	3	7	9	15	24	28	31	34	81
33	Pagelaran	13	17	33	44	50	64	66	75	88	194

Tabel 3. Datascale pasien Covid-19

	Bulan 10	Bulan 11	Bulan 12	Bulan 1	Bulan 2	Bulan 3	Bulan 4	Bulan 5	Bulan 6	Bulan 7
1	-0.31358	-0.37749	-0.46664	-0.51863	-0.54404	-0.48875	-0.31495	-0.18494	0.102673	-0.19127
2	-0.60508	-0.62873	-0.62289	-0.70703	-0.65463	-0.65285	-0.62441	-0.63952	-0.61017	-0.57392
3	-0.5322	-0.52405	-0.53609	-0.62007	-0.5809	-0.52391	-0.47502	-0.49444	-0.49577	-0.57786
4	-0.50791	-0.50311	-0.57081	-0.59109	-0.55633	-0.61769	-0.63508	-0.62985	-0.60137	-0.43979
5	-0.21641	-0.16813	-0.0847	-0.19981	-0.12623	-0.09022	-0.05885	-0.08822	-0.02934	-0.02164
6	-0.62937	-0.64967	-0.69234	-0.76499	-0.76523	-0.78179	-0.81648	-0.83296	-0.79499	-0.99602
7	-0.50791	-0.48217	-0.41456	-0.37371	-0.43344	-0.46531	-0.45367	-0.44608	-0.46056	-0.59364
8	-0.36216	-0.33562	-0.43192	-0.34473	-0.34742	-0.45359	-0.47502	-0.51378	-0.50457	-0.47924
9	-0.28929	-0.27281	-0.15414	0.220451	0.266992	0.237981	0.197249	0.230954	0.20828	0.47147
10	-0.21641	-0.14719	-0.22359	-0.2143	-0.13852	-0.10194	-0.13355	-0.1656	-0.16134	-0.47924
11	-0.45933	-0.48217	-0.51873	-0.62007	-0.60548	-0.50047	-0.52837	-0.51378	-0.51337	-0.51869
12	-0.55649	-0.48217	-0.41456	-0.34473	-0.26141	-0.18399	-0.14422	-0.19461	-0.60137	-0.34512
13	0.949572	0.941509	0.870155	0.829108	1.102598	1.257749	1.456413	1.565679	1.73958	1.950791
14	0.196539	0.104048	-0.01526	0.061041	-0.02793	-0.05506	-0.05885	-0.11724	-0.16134	-0.32934
15	-0.33787	-0.31468	-0.34512	-0.43168	-0.4826	-0.52391	-0.5177	-0.50411	-0.49577	-0.73566
16	-0.24071	-0.23094	-0.22359	-0.30125	-0.32285	-0.38326	-0.40032	-0.35903	-0.34616	-0.56997
17	-0.21641	-0.25187	-0.36248	-0.34473	-0.372	-0.43014	-0.46434	-0.43641	-0.45176	-0.53447
18	1.726897	1.653352	1.963899	2.002946	2.048798	1.867265	1.840564	1.904196	1.818785	1.674651
19	0.536619	0.480906	0.679184	0.611731	0.512758	0.495853	0.45335	0.501768	0.481098	0.542477
20	-0.5322	-0.56592	-0.57081	-0.59109	-0.5809	-0.52391	-0.443	-0.44608	-0.44296	-0.43979
21	0.172248	0.208731	0.088909	0.090025	-0.01564	-0.05506	-0.08019	-0.07855	-0.04694	0.033591
22	0.366579	0.376223	0.384047	0.394353	0.439028	0.495853	0.45335	0.385704	0.296286	0.668713
23	0.706658	0.794954	0.679184	0.988518	1.016579	1.093648	1.125616	1.082083	0.956329	0.727886
24	4.204618	4.333227	4.238192	3.901375	3.79374	3.777864	3.707967	3.645142	3.631703	3.647079
25	1.994102	1.778971	1.703484	2.060913	2.073375	2.101694	2.118007	2.107306	2.144406	1.887673
26	-0.36216	-0.29375	-0.18887	-0.14184	-0.22454	-0.26604	-0.33629	-0.38805	-0.44296	-0.64887
27	-0.43504	-0.46124	-0.484	-0.46066	-0.53175	-0.5708	-0.62441	-0.64919	-0.54857	-0.09659
28	-0.72653	-0.69154	-0.67498	-0.73601	-0.77752	-0.82867	-0.85917	-0.87165	-0.88299	-0.12815
29	-0.60508	-0.58686	-0.62289	-0.70703	-0.74065	-0.79351	-0.81648	-0.82329	-0.82139	-0.92501
30	-0.65366	-0.6706	-0.65761	-0.67804	-0.70379	-0.77006	-0.74179	-0.70722	-0.65418	-0.21888
31	-0.43504	-0.46124	-0.46664	-0.53312	-0.50717	-0.38326	-0.40032	-0.40739	-0.42536	-0.51474
32	-0.67795	-0.69154	-0.65761	-0.72152	-0.6915	-0.67629	-0.67776	-0.67821	-0.68058	-0.84611
33	-0.43504	-0.39843	-0.20623	-0.2143	-0.26141	-0.20743	-0.27227	-0.25264	-0.20535	-0.40034

4.2. Datascale

Pada data numerik pasien Covid-19 diatas perlu adanya penskalaan data karena satuan dari data numerik masih tidak beraturan. Maka perlu adanya penskalaan data yang fungsinya untuk menyamakan satuan data dari data numerik pasien Covid-19 diatas. Penskalaan data Covid-19 menggunakan rumus dasar sebagai berikut:

$$data = \frac{data\ numerik - mean(data\ numerik)}{standart\ deviasi(data\ numerik)} \quad [13]$$

Syntax Rstudio dari *data scale* adalah sebagai berikut:

```
datascale <- scale(datanumerik)
```

Sehingga hasil data numerik setelah dilakukan penskalaan yang dapat dilihat pada tabel 3.

Pada tabel 3 penskalaan data pasien Covid-19, data yang diambil adalah data dari keseluruhan pasien pada masing-masing Kecamatan. Tujuan dari penskalaan data pasien Covid-19 adalah untuk menyamakan satuan dari masing-masing data pasien Covid-19.

4.3. Euclidean Distance

Tahapan pertama dari perhitungan *Clustering* dengan menggunakan metode K-Means *Euclidean Distance* adalah mengambil data Covid-19 yang sudah di lakukan penskalaan. Tahap kedua yang digunakan untuk menghitung *Euclidean Distance* adalah menentukan nilai klaster. Dalam penelitian ini penulis menggunakan klaster atau $k = 3$.

Tahap selanjutnya adalah mengambil data yang sudah dilakukan penskalaan kemudian dimasukkan ke dalam perhitungan jarak *Euclidean Distance* dengan rumus dasar seperti dibawah ini:

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad [6]$$

Perhitungan k-means *Clustering* menggunakan Software Rstudio dapat dilakukan dengan Syntax seperti dibawah ini:

```
clseu <- Kmeans(datascale, method
  = "euclidean", centers
  = 3, iter.max = 20, nstart
  = 25)
```

Hasil dari syntax di atas yaitu:

Tabel 4. Hasil *Clustering* Kecamatan dengan *Euclidean Distance*

Kecamatan	Cluster Kecamatan	Within Cluster sum of squares	Size
Donomulyo	3	0	1
Pagak	3	2.468845025	4
Bantur	3	0.131145605	28
Sumbermanjing Wetan	3	0	1
Dampit	3	2.468845025	4
Ampelgading	3	0.131145605	28
Poncokusumo	3	0	1
Wajak	3	2.468845025	4
Turen	3	0.131145605	28
Gondanglegi	3	0	1
Kalipare	3	2.468845025	4
Sumberpucung	3	0.131145605	28
Kepanjen	2	0	1
Bululawang	3	2.468845025	4
Tajinan	3	0.131145605	28
Tumpang	3	0	1
Jabung	3	2.468845025	4
Pakis	2	0.131145605	28
Pakisaji	3	0	1
Ngajum	3	2.468845025	4
Wagir	3	0.131145605	28
Dau	3	0	1
Karangploso	2	2.468845025	4
Singosari	1	0.131145605	28
Lawang	2	0	1
Pujon	3	2.468845025	4
Ngantang	3	0.131145605	28
Kasembon	3	0	1
Gedangan	3	2.468845025	4
Tirtoyudo	3	0.131145605	28
Kromengan	3	0	1
Wonosari	3	2.468845025	4
Pagelaran	3	0.131145605	28

Dari Tabel 4 hasil perhitungan *K-Means Euclidean Distance* menunjukkan bahwa pada masing-masing Kecamatan di Kabupaten Malang menempati *Cluster* pada kolom *Cluster Euclidean*.

4.4. Manhattan Distance

Tahap pertama dalam perhitungan K-Means *Manhattan Distance* adalah dengan menentukan nilai dataset yang akan dipakai dalam perhitungan. Dataset yang digunakan adalah data dari pasien Covid-19 pada bulan oktober 2020 – Juli 2021 yang sudah dilakukan penskalaan.

Tahap kedua adalah menentukan nilai klaster atau k yang fungsinya digunakan untuk menghitung persamaan jarak pada metode K-Means yaitu *Manhattan Distance*. pada penelitian ini penulis menggunakan nilai $k = 3$.

Tahap ketiga adalah melakukan perhitungan dengan rumus K-Means *Manhattan Distance* yaitu:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad [6]$$

Pada penelitian ini penulis menggunakan Rstudio untuk melakukan perhitungan K-Means *Manhattan Distance*. Tahapan dari perhitungan K-Means *Manhattan Distance* pada Rstudio adalah dengan memasukkan syntax sebagai berikut:

```
clsman <- kmeans(datascale, method
  = "manhattan", centers
  = 3, iter.max = 20, nstart
  = 25)
```

Hasil dari syntax di atas yaitu:

Tabel 1. Hasil *Clustering* Kecamatan dengan *Manhattan Distance*

Kecamatan	Cluster Manhattan	Within Cluster Sum of Squares	Size
Donomulyo	2	8.753864	5
Pagak	2	2.545197	25
Bantur	2	33.80413	3
Sumbermanjing Wetan	2	8.753864	5
Dampit	2	2.545197	25
Ampelgading	2	33.80413	3
Poncokusumo	2	8.753864	5
Wajak	2	2.545197	25
Turen	1	33.80413	3
Gondanglegi	2	8.753864	5
Kalipare	2	2.545197	25
Sumberpucung	2	33.80413	3
Kepanjen	1	8.753864	5
Bululawang	2	2.545197	25
Tajinan	2	33.80413	3
Tumpang	2	8.753864	5
Jabung	2	2.545197	25
Pakis	3	33.80413	3
Pakisaji	1	8.753864	5
Ngajum	2	2.545197	25
Wagir	2	33.80413	3
Dau	1	8.753864	5
Karangploso	1	2.545197	25
Singosari	3	33.80413	3
Lawang	3	8.753864	5
Pujon	2	2.545197	25
Ngantang	2	33.80413	3
Kasembon	2	8.753864	5
Gedangan	2	2.545197	25
Tirtoyudo	2	33.80413	3
Kromengan	2	8.753864	5
Wonosari	2	2.545197	25
Pagelaran	2	33.80413	3

Dari Tabel 5 hasil perhitungan K-Means *Manhattan Distance* menunjukkan bahwa pada masing-masing Kecamatan di Kabupaten Malang menempati *Cluster* pada kolom *Cluster Manhattan*.

4.5. Silhouette Coefficient dan Standar Deviasi

Berikut adalah hasil perbandingan metode K-Means *Euclidean Distance* dan *Manhattan Distance* menggunakan metode *Silhouette Coefficient* dan Standar Deviasi yaitu:

Tabel 2. Hasil Perbandingan

Metode	Silhouette Coefficient	Standart Deviasi
K-Means <i>Euclidean Distance</i>	0,71	0,46466002
K-Means <i>Manhattan Distance</i>	0,64	0,4961977

Dari data tabel 6 diatas menunjukkan bahwa Metode K-Means dengan *Euclidean Distance* memperoleh nilai *Silhouette Coefficient* yang lebih baik dari metode K-Means *Manhattan Distance* dengan perbandingan $0,71 > 0,64$.

Sedangkan nilai Standar Deviasi dari kedua Metode K-Means menunjukkan bahwa *Euclidean Distance* mempunyai nilai yang lebih rendah dibandingkan *Manhattan Distance* dengan perbandingan $0,46466002 < 0,4961977$. Semakin rendah nilai Standar Deviasi maka semakin mendekati rata-rata atau semakin baik.

Dapat disimpulkan bahwa metode K-Means dengan model *Euclidean Distance* lebih baik dibandingkan Metode K-Means dengan model *Manhattan Distance*.

5. KESIMPULAN DAN SARAN

Pada bagian akhir dari penelitian ini penulis dapat menyimpulkan beberapa kesimpulan diantaranya adalah sebagai berikut: Metode K-Means dengan model *Euclidean Distance* dan *Manhattan Distance* dapat digunakan dalam penentuan Zonasi Covid-19 di Kabupaten Malang. Hasil dari perbandingan perhitungan jarak *Euclidean* dan *Manhattan Distance* menggunakan *Silhouette Coefficient* menunjukkan bahwa *Euclidean Distance* lebih unggul dibandingkan *Manhattan Distance* dengan hasil perbandingan $0,71 > 0,64$. Sedangkan untuk Hasil dari perhitungan Standar Deviasi menunjukkan bahwa *Euclidean Distance* lebih unggul dibandingkan *Manhattan Distance* dengan hasil perbandingan $0,46466002 < 0,4961977$. Saran yang dapat disampaikan oleh penulis untuk penelitian selanjutnya agar dapat dikembangkan oleh peneliti selanjutnya adalah sebagai berikut: pada penelitian selanjutnya diharapkan dapat mengimplementasikan metode K-Means dengan menggunakan metode pengukuran kualitas hasil *clustering* yang lain dan dapat membandingkan model perhitungan K-Means dengan model perhitungan jarak yang lain.

DAFTAR PUSTAKA

- [1] M. Metisen and H. L. Sari, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhillah," *Jurnal Media Infotama*, p. 112, 2015
- [2] A. and R. Adrian, "penerapan metode K-means untuk clustering mahasiswa berdasarkan nilai akademik dengan Weka interface studi kasus pada jurusan teknik informatika umm malang," *JURNAL ILMIAH SEMESTA TEKNIKA*, pp. 77-82, 2015

- [3] A. Susilo, "Coronavirus Disease 20219: Tinjauan Literatur Terkini," *Jurnal penyakit dalam indonesia*, pp. 45-46, 2020
- [4] A. Wanto, Data Mining: Algoritma dan Implementasi, yayasan kita menulis, 2020
- [5] A. H. Fitri and E. Prasetyo, "penerapan metode clustering dengan algoritma K-Means pada pengelompokan data calon mahasiswa baru di universitas muhammadiyah Yogyakarta," *SEMESTA TEKNIKA*, pp. 60-64, 2018
- [6] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, pp. 20-24, 2019
- [7] R. I. Fajriah, H. Sutisna and B. K. Simpony, "Perbandingan Distance Space Manhattan Distance dengan Euclidean Distance pada K-Means Clustering dalam menentukan promosi," *IJCIT (Indonesian Journal on Computer and Information Technology)*, pp. 37-49, 2019
- [8] R. Awasthi, A. K. Tiwari and S. Pathak, "Empirical Evaluation on K Means Clustering with Effect of Distance Function for Bank Dataset," (*IJITR*) *INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH*, pp. 233-235, 2013
- [9] A. Aziz, D. P. Pamungkas and A. B. Setiawan, "Analisa perbandingan Algoritma Euclidean Distance dan Manhattan Distance dalam identifikasi wajah," *Seminar Nasional Inovasi Teknologi*, pp. 219-224, 2021
- [10] S. Paembonan and H. Abduh, "Penerapan metode Silhouette Coefficient untuk evaluasi clustering obat," *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, pp. 48-54, 2021
- [11] R. Ananda and M. Fadhli, Statistik pendidikan teori dan praktik dalam pendidikan, Medan: CV. Widya Puspita, 2018, p. 57
- [12] Sugiyono, "Metode Penelitian Kuantitatif, Kualitatif dan R&D," *alfabeta*, 2012
- [13] Kukuh, "machine learning-ID," 10 febuari 2018. [Online]. Available: <https://medium.com/machine-learning-id/melakukan-feature-scaling-pada-dataset-229531bb08de>