



IMPLEMENTASI *K-MEANS CLUSTERING* DALAM PEMETAAN WILAYAH RAWAN PENYAKIT LEPTOSPIROSIS DI JAWA TIMUR

SALSABILA¹, MOH.HAFIYUSHOLEH², HANI KHAULASARI³, DIAN CANDRA R.N.⁴, YUNIARTI D.N.⁵

¹Program Studi Matematika UIN Sunan Ampel Surabaya

^{2,3,4}UIN Sunan Ampel Surabaya

⁵Dinas Kesehatan Provinsi Jawa Timur

*penulis korespondensi: hafiyusholeh@uinsa.ac.id

ABSTRAK

Leptospirosis merupakan penyakit endemis yang banyak ditemukan pada wilayah dengan tingkat curah hujan tinggi. Bakteri *Leptospira sp.* merupakan bakteri penyebab Leptospirosis yang dapat menginfeksi hewan dan manusia secara langsung maupun tidak langsung. Kasus Leptospirosis termasuk kasus terabaikan dan terus meningkat setiap tahunnya, khususnya pada Provinsi Jawa Timur yaitu ditemukan 273 kasus di Tahun 2020 menjadi 606 kasus di Tahun 2023. Penelitian ini memiliki tujuan untuk melakukan pemetaan wilayah rawan Leptospirosis di Jawa Timur menggunakan *K-Means Clustering*. Data yang digunakan terdiri dari jumlah kasus Leptospirosis, jumlah kematian akibat Leptospirosis, CFR, jumlah penduduk, jumlah petani, jumlah kasus banjir, jumlah kasus tanah longsor, luas wilayah, dan luas wilayah tegal/kebun. Hasil penelitian didapatkan 5 kategori *cluster* yaitu C1 dengan kategori sangat aman terdiri dari 4 Kab/Kota, C2 kategori aman terdiri dari 7 Kabupaten dan Kota, C3 kategori cukup aman terdiri dari 13 Kabupaten dan Kota, C4 kategori sangat rawan terdiri dari 11 Kabupaten dan Kota, dan C5 kategori rawan terdiri dari 3 Kabupaten dan Kota di Provinsi Jawa Timur dengan hasil akurasi sebesar 0.65 menggunakan *Silhouette Coefficient*.

Kata Kunci: Leptospirosis, *K-Means Clustering*, Pemetaan Wilayah

ABSTRACT

Leptospirosis is an endemic disease that is commonly found in areas with high levels of rainfall. Leptospira sp. is a bacterium that causes Leptospirosis which can infect animals and humans directly or indirectly. Leptospirosis cases are neglected cases and continue to increase every year, especially in East Java Province where 273 cases were found in 2020 to 606 cases in 2023. This study aims to map Leptospirosis-prone areas in East Java using K-Means Clustering. The data used consists of the number of Leptospirosis cases, the number of deaths due to Leptospirosis, CFR, population, number of farmers, number of flood cases, number of landslide cases, area, and area of farmland. So that the research results obtained 5 cluster categories, namely C1 with a very safe category consisting of 4 districts/cities, C2 safe category consisting of 7 districts/cities, C3 fairly safe category consisting of 13 districts/cities, C4 very vulnerable category consisting of 11 districts/cities, and C5 vulnerable category consisting of 3 districts/cities in East Java Province with an accuracy result of 0.65 using Silhouette Coefficient.

Keywords: *Leptospirosis, K-Means Clustering, Mapping Territory*

1 Pendahuluan

Leptospirosis merupakan penyakit endemis yang banyak ditemukan pada wilayah dengan tingkat curah hujan tinggi [1]. Penyakit ini dapat dikatakan sebagai penyakit musiman karena penyakit ini akan meningkat dengan adanya peningkatan curah hujan dan jumlah daerah banjir. Sebagian

wilayah di Indonesia termasuk ke dalam daerah endemis Leptospirosis, namun tingginya biaya pemeriksaan laboratorium dan minimumnya metode diagnosis yang tepat menyebabkan tidak seluruhnya kasus Leptospirosis dilaporkan [2]. Selain itu penyakit ini juga termasuk kedalam salah satu masalah kesehatan yang tidak diperhatikan [3].

Bakteri *Leptospira sp.* merupakan bakteri penyebab Leptospirosis yang dapat menginfeksi hewan atau manusia secara langsung yaitu melalui darah, urin, dan cairan tubuh lainnya ataupun tidak langsung yaitu melalui genangan air, sungai, selokan, saluran air, dan lumpur yang terdapat urin hewan yang mengandung bakteri *Leptospira*. Gejala awal yang ditimbulkan adalah seperti flu, namun jika tidak diketahui dan diatasi dengan segera maka akan menyebabkan infeksi pada hati dan ginjal yang pada akhirnya menyebabkan kematian [4]. Sejak tahun 2021 di Indonesia telah ditemukan 734 kasus Leptospirosis dan terus meningkat pada tahun 2022 yaitu telah dilaporkan 1.419 kasus Leptospirosis, 139 kematian akibat Leptospirosis, dan CFR (*Case Fatality Rate*) sebesar 9,8%, hal ini menandakan angka CFR telah melebihi ambang batas indikator pemberantas penyakit zoonosis yaitu $>7\%$ dan Provinsi Jawa Timur menjadi penyumbang kasus Leptospirosis terbesar di Indonesia yaitu sebesar 28,3% [5].

Kasus Leptospirosis di Jawa Timur berdasarkan laporan tahunan kasus Leptospirosis Dinas Kesehatan Jawa Timur, sejak Tahun 2020-2022 terus mengalami peningkatan, yaitu pada tahun 2020 terdapat 273 kasus dengan 21 kasus kematian, Tahun 2021 terdapat 312 kasus dengan 29 kasus kematian, dan Tahun 2022 terdapat 606 kasus dengan 22 kasus kematian yang dilaporkan.

Meningkatnya jumlah kasus Leptospirosis dan kematian akibat Leptospirosis dari tahun ke tahun khususnya di Jawa Timur menunjukkan terdapat berbagai permasalahan yang belum terselesaikan dengan optimal. Salah satu permasalahan tersebut diantaranya adalah kurangnya kesadaran dan kewaspadaan tenaga kesehatan dan minimnya informasi serta pengetahuan masyarakat umum terkait penyakit Leptospirosis yang menjadikan tantangan yang menyebabkan kasus Leptospirosis terus meningkat setiap tahunnya [6]. Hal ini ditunjukkan dengan tidak masuknya gejala Leptospirosis sebagai diagnosa banding pada deteksi awal pasien yang memiliki gejala yang sama dengan penyakit Leptospirosis sehingga menyebabkan adanya kasus yang tidak terdiagnosis dengan cepat dan tepat yang mengakibatkan banyaknya pasien yang terlambat untuk ditangani sehingga peluang kesembuhan kecil [7]. Selain itu tidak optimalnya pencatatan dan pelaporan kasus Leptospirosis menjadi permasalahan utama dalam tindakan untuk menanggulangi penyakit ini, sebab jika tidak terdapat kasus yang dilaporkan maka akan sulit untuk segera diatasi dan dilakukan penanganan yang tepat dan cepat [8]. Tidak adanya pencatatan dan pelaporan kasus Leptospirosis mengakibatkan beberapa wilayah tidak terpetakan dengan baik atau kurangnya informasi terkait wilayah mana saja yang memiliki kasus tertinggi dan diperlukannya perhatian khusus, oleh karena itu perlunya dilakukan pemetaan wilayah rawan Leptospirosis.

Pemetaan wilayah rawan penyakit Leptospirosis perlu dilakukan untuk mempermudah pemerintah daerah dan pihak terkait dalam mengatasi wilayah agar dapat dilakukan mitigasi terhadap penyebaran penyakit Leptospirosis sehingga dapat dilakukanantisipasi, mitigasi, dan meminimalisir resikonya. Penelitian ini juga memiliki tujuan untuk meningkatkan kesadaran dan kewaspadaan masyarakat umum terhadap kasus Leptospirosis, sehingga diharapkan kasus Leptospirosis di tahun selanjutnya mengalami penurunan. Pemetaan wilayah yang dapat dilakukan adalah dengan menggunakan analisis *clustering*. Terdapat berbagai metode *clustering* untuk pemetaan wilayah diantaranya adalah metode *Fuzzy C-Means* [9], *Hybrid Hierarchical Clustering* [10], dan metode *K-Means Clustering* [11].

K-Means Clustering termasuk metode *unsupervised learning* yang banyak digunakan untuk membagi data ke dalam beberapa kelompok atau *cluster* [12]. Metode ini memiliki beberapa keunggulan yaitu mudah dan sederhana untuk implementasikan, lebih efektif dan efisien dalam

pengelompokan data, hasilnya mudah diinterpretasikan, dan memiliki tingkat akurasi yang cukup tinggi [13].

Beberapa penelitian terkait dengan *K-Means Clustering* antara lain penelitian yang dilakukan oleh A. Rizal, dkk yaitu membahas mengenai perbandingan metode *K-Means Clustering*, *Fuzzy C-Means* dan *Probabilistic Distance Clustering* dan disimpulkan bahwa metode *K-Means Clustering* memiliki nilai akurasi yang lebih tinggi dibandingkan metode lainnya yaitu sebesar 0,5137, metode *Fuzzy C-Means* (FCM) sebesar 0,5132, dan metode *Probabilistic Distance Clustering* (PDC) sebesar 0,5114 [14]. Penelitian lainnya yang dilakukan oleh Putri dan Dwidayati terkait perbandingan antara *K-Means Clustering* dengan *Fuzzy C-Means* dengan hasil penelitian menunjukkan bahwa pengujian menggunakan Dunn-Index dan Davies Bouldin-Index bahwa *clustering K-Means* memiliki hasil akurasi yang lebih besar yaitu 1,165219 dibandingkan dengan *Fuzzy C-Means*[15]. Penelitiannya lainnya dilakukan oleh Nurvita terkait pemetaan epidemiologi Leptospirosis berbasis system informasi geografis tahun 2021-2022 juga telah dilakukan dengan menghasilkan hasil akhir bahwa morbiditas Leptospirosis tahun 2021-2022 di Puskesmas Kedungmundu mengalami penurunan, namun mengalami peningkatan pada mortalitas Leptospirosis pada tahun 2021-2022 [16]. Penelitian selanjutnya dilakukan oleh Ganinov dan Huda terkait penerapan system informasi geografis faktor risiko penyakit Leptospirosis yang didapatkan hasil bahwa pemetaan kasus Leptospirosis di Kabupaten Bantul terdapat 26 kasus Leptospirosis dan diketahui terdapat hubungan yang signifikan terkait keberadaan tikus dengan penyakit Leptospirosis di Kabupaten Bantul [17].

Berdasarkan uraian diatas, maka penting dilakukan pemetaan terhadap wilayah rawan Leptospirosis di wilayah Jawa Timur. Perbedaan penelitian ini dengan penelitian sebelumnya adalah penelitian ini dilakukan di Provinsi Jawa Timur Tahun 2023 yang belum pernah dilakukan sebelumnya. Pembaharuan penelitian ini yaitu menggunakan metode *K-Means Clustering* dan menambahkan variabel pendukung lainnya seperti data bencana alam dan luas wilayah yang tidak digunakan oleh penelitian sebelumnya [18]. Penelitian ini diharapkan dapat memberikan informasi dengan cakupan yang lebih luas, mengingat kasus Leptospirosis di Jawa Timur terus mengalami peningkatan setiap tahunnya. Selain itu penelitian ini juga diharapkan dapat membantu pihak pemerintah khususnya tenaga kesehatan baik ditingkat puskesmas, rumah sakit, dan masyarakat umum di Indonesia khususnya Provinsi Jawa Timur untuk meningkatkan kesadaran, kewaspadaan, dan kepedulian terhadap kasus Leptospirosis. Sehingga dapat dilakukan upaya pencegahan dan pengendalian yang cepat dan tepat sasaran dan diharapkan pula pada tahun selanjutnya kasus Leptospirosis di Jawa Timur dapat mengalami penurunan.

2 Tinjauan Pustaka

2.1 Leptospirosis

Leptospirosis adalah penyakit zoonosis menular yang disebabkan oleh bakteri *Leptospira sp.* yang berberbentuk spiral dan dapat menyerang hewan dan manusia [19]. Pada umumnya di Indonesia bakteri *Leptospira* ditemukan pada urine tikus, namun bakteri tersebut juga dapat ditemukan pada urine kucing, anjing, dan hewan ternak seperti sapi, kambing, dan babi. Penularan bakteri ini dapat terjadi secara langsung dan tidak langsung yaitu dengan cara melalui darah, urin dan cairan tubuh lainnya yang terdapat bakteri *Leptospira* yang masuk melalui luka pada kulit manusia dan penularan dapat terjadi secara tidak langsung melalui perantara berupa air seperti pada genangan air, sungai, danau, selokan, saluran air, maupun lumpur yang telah terkontaminasi oleh bakteri *Leptospira* [20]. Lama waktu yang dibutuhkan sejak tubuh telah terinfeksi bakteri *Leptospira* hingga muncul gejala adalah sekitar 7-10 hari [21]. Adapun gejala yang timbul setelah 7-10 hari terinfeksi bakteri adalah munculnya demam, diare, nyeri otot pada betis, mata merah, ruam pada kulit, gangguan pada saat pembuangan urine, dan munculnya

penyakit kuning atau ikterus. Namun apabila terjadi keterlambatan diagnosa dan penanganan terhadap penyakit ini, maka akan menimbulkan gangguan pada fungsi ginjal yang menyebabkan kematian [22].

Risiko tinggi manusia untuk terpapar penyakit ini adalah adanya faktor pekerjaan dan lingkungan sekitar yang tidak baik. Faktor pekerjaan yang menyebabkan manusia memiliki risiko tinggi adalah pekerjaan sebagai petani, peternak, petugas kebersihan, pekerja perkebunan dan profesi lainnya yang berkaitan dengan hewan dan lingkungan bebas [23]. Lingkungan yang buruk juga menyebabkan manusia memiliki faktor risiko untuk tertular, diantaranya adalah adanya kepadatan penduduk, banyaknya populasi tikus di sekitar rumah, sanitasi lingkungan yang buruk, selokan yang tidak terawat, banyaknya genangan air dan lain sebagainya [24]. Wilayah rawan banjir dan tanah longsor juga termasuk faktor penularan penyakit *Leptospirosis*, sebab air dan tanah merupakan tempat yang mudah terkontaminasi oleh bakteri *Leptospira* [25].

2.2 K-Means Clustering

Metode *K-Means Clustering* termasuk kedalam teknik *clustering* non-hirarki yang digunakan untuk mengelompokkan objek dengan menentukan jumlah *cluster* nya terlebih dahulu [26]. *K-Means* bertujuan untuk mengelompokkan data dalam bentuk satu atau lebih jumlah *cluster*, data yang memiliki karakteristik yang sama akan masuk kedalam *cluster* yang sama, begitu pula sebaliknya data yang memiliki karakteristik yang berbeda akan masuk kedalam *cluster* yang lainnya sehingga tujuan dari metode ini adalah untuk meminimumkan variasi dalam suatu *cluster* dan memaksimumkan variasi antar *cluster* [27]. Berikut merupakan tahapan-tahapan pada metode *K-Means Clustering* [28]:

- 1) Menentukan jumlah *cluster* k yang akan diuji,
- 2) Menentukan titik pusat atau centroid awal dengan cara memilih secara acak yang disesuaikan dengan jumlah cluster yang telah ditentukan sebelumnya. Misalkan *centroid* ini adalah $\mu_1, \mu_2, \dots, \mu_k$
- 3) Menghitung jarak terdekat data dengan titik pusat data (*centroid*) menggunakan cara *Euclidean Distance* dengan rumus:

$$d_{(x_i, \mu_j)} = \sqrt{\sum_{l=1}^n (x_{il} - \mu_{jl})^2}. \quad (1)$$

Keterangan:

$d_{(x, \mu)}$ = Jarak data x menuju titik pusat cluster μ

x_i = Data x pada pengamatan sebanyak- i

μ_i = Titik pusat (*centroid*) ke- j

n = Total atau banyaknya pengamatan

dengan data x_i akan diberikan ke kluster C_j yang memiliki jarak minimum, yaitu

$$C_j = \{x_i: \forall j, d_{(x_i, \mu_j)} \leq d_{(x_i, \mu_m)}, m \neq j\}.$$

- 4) Menghitung kembali pusat *cluster* dengan cara menghitung nilai rata-rata (μ) dari seluruh data pada setiap *cluster*.

$$\mu_j = \frac{1}{N_j} \sum_{x_i \in C_j} x_i$$

- 5) Mengulangi perhitungan kembali pada setiap objek dengan menggunakan nilai centroid baru, apabila tidak terdapat perbedaan atau perubahan pada anggota *cluster* maka tahapan analisis *clustering* selesai.

2.3 *Silhouette Coefficient*

Silhouette Coefficient merupakan salah satu bentuk evaluasi dari analisis *clustering* data yang bertujuan untuk mengetahui atau mengukur kualitas pengelompokan yang telah terbentuk [29]. Metode ini digunakan untuk menghitung kedekatan jarak dalam satu *cluster* dengan *cluster* lainnya. *Silhouette Coefficient* memiliki rentang nilai antara -1 sampai dengan 1, dimana hasil perhitungan menghasilkan angka mendekati -1 menunjukkan bahwa data tersebut dikelompokkan dengan tidak baik, begitu pula sebaliknya bahwa apabila hasil yang ditunjukkan semakin mendekati nilai 1 maka hasil pengelompokan semakin baik [30]. Berikut persamaan *Silhouette Coefficient* yaitu [31]:

$$S_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad (2)$$

Keterangan:

i = objek yang akan diteliti

a_i = Nilai rata-rata jarak antar anggota dalam suatu *cluster*

b_i = Nilai terkecil rata-rata jarak dari objek ke- i dengan objek yang berada di *cluster* lainnya

S_i = Nilai *Silhouette Coefficient* objek ke- i dalam suatu *cluster*

Berikut merupakan penentuan kualitas pengelompokan berdasarkan nilai yang dihasilkan menggunakan metode *Silhouette Coefficient* yaitu sebagai berikut [32]:

Tabel 1. Interpretasi Nilai *Silhouette Coefficient*

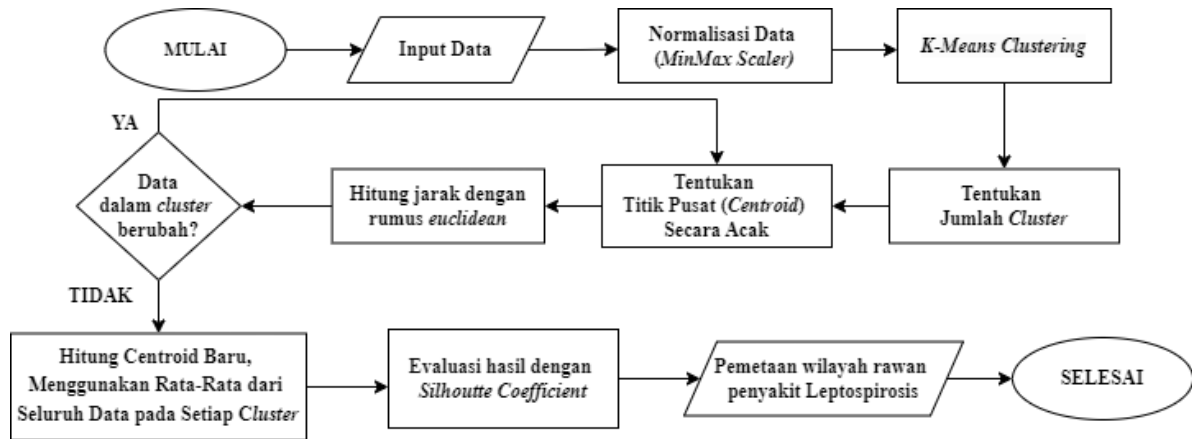
<i>Silhouette Coefficient</i>	Akurasi
$\geq 0.7-1$	Sangat Baik
0.5-0.7	Baik
$0.25 \leq 0.5$	Buruk
≤ 0.25	Sangat Buruk

3 Hasil dan Pembahasan

3.1 Tahapan Penelitian

Data yang digunakan bersumber dari Dinas Kesehatan Jawa Timur, website resmi BNPB (Badan Nasional Penanggulangan Bencana) [33], dan website resmi BPS (Badan Pusat Statistik) [34]. Penelitian ini menggunakan beberapa variable yang terdiri dari jumlah kasus (X1), jumlah kematian (X2), CFR (X3), jumlah penduduk (X4), jumlah petani (X5), jumlah kasus banjir (X6), jumlah kasus tanah longsor (X7), luas wilayah (X8), dan luas wilayah tegal/kebun (X9). Data yang digunakan merupakan data Provinsi Jawa Timur tahun 2023.

Tahapan dalam penelitian ini adalah pertama, mengumpulkan dan input data yang dibutuhkan terkait penelitian ini dengan berbagai macam sumber dengan fokus penelitian pada wilayah Provinsi Jawa Timur pada tahun 2023. Kedua, melakukan tahapan normalisasi data menggunakan metode *Min-Max Scaler* sebelum dilakukan analisis data. Ketiga, melakukan tahapan analisis data dengan metode *k-means clustering* dengan langkah awal yaitu menentukan jumlah *cluster* yang akan digunakan atau diuji, menentukan titik pusat atau nilai *centroid* awal dengan memilih secara acak, menghitung jarak terdekat data dengan nilai *centroid*, menggunakan rumus jarak *Euclidean*, menghitung kembali pusat *cluster* dengan cara menghitung nilai rata-rata dari seluruh data pada setiap *cluster*, dan selanjutnya adalah melakukan tahap evaluasi dengan menggunakan metode *Silhouette Coefficient*. Terakhir, melakukan visualisasi hasil penelitian dengan membentuk suatu peta mengenai wilayah rawan penyakit *Leptospirosis*. Berikut merupakan diagram alir pada Gambar 1.



Gambar 1. Diagram Alir K-Means Clustering

3.2 Deskripsi Data

Penelitian ini menggunakan 38 data dan 9 variabel yang berkaitan dengan kasus Leptospirosis yaitu terdiri dari jumlah kasus Leptospirosis, jumlah kematian akibat Leptospirosis, *case fatality rate* (CFR), jumlah penduduk, jumlah petani, jumlah kasus banjir, jumlah kasus tanah longsor, luas wilayah, dan luas wilayah tegal/kebun pada tahun 2023 di Provinsi Jawa Timur. Sampel data Leptospirosis Tahun 2023 di Provinsi Jawa Timur dapat dilihat pada Tabel 2.

Tabel 2. Sampel data Leptospirosis Tahun 2023 di Provinsi Jawa Timur

No.	Kabupaten/ Kota	Jumlah Kasus	Jumlah Kematian	...	Luas Wilayah	Luas Wilayah Tegal/Kebun
1.	Pacitan	918	6	...	1.433.59	470.82
2.	Ponorogo	17	1	...	1.418.62	392.94
...
37.	Pasuruan	12	0	...	1.493.29	320.50
38.	Sampang	22	0	...	1.228.25	778.38

3.3 Normalisasi Data

Selanjutnya, dilakukan pengolahan data yaitu diketahui sampel data pada Tabel 2 akan dilakukan tahap normalisasi dengan menggunakan metode *Min-Max Scaler*. Tujuan normalisasi data yaitu untuk mengubah rentang nilai data yang ada agar nilai data tersebut berada dalam rentang nilai yang sama. Langkah yang dilakukan adalah dengan menentukan nilai minimum (x_{min}) dan maksimum (x_{max}) untuk setiap fitur pada dataset kemudian hitung nilai transformasi x_{scaled} untuk setiap nilai x dalam fitur dengan menggunakan rumus

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Berikut ditampilkan Tabel 3 hasil normalisasi menggunakan metode *Min-Max Scaler*.

Tabel 3. Normalisasi Data menggunakan Min-Max Scaler

No.	Kabupaten/ Kota	Jumlah Kasus	Jumlah Kematian	...	Luas Wilayah	Luas Wilayah Tegal/Kebun
1.	Pacitan	1.00	1.00	...	3.95	4.00
2.	Ponorogo	1.85	1.66	...	3.91	3.34
...
37.	Pasuruan	2.50	6.66	...	4.77	3.87

38.	Sampang	0.00	0.00	...	3.58	5.31
-----	---------	------	------	-----	------	------

Setelah dilakukan normalisasi menggunakan metode *Min-Max Scaler* didapatkan hasil yaitu data yang pada awalnya memiliki rentang yang berbeda-beda disetiap variabelnya, saat ini data telah berada didalam rentang yang sama, sehingga data yang telah dilakukan normalisasi dapat diolah dengan baik.

3.4 Perhitungan *K-Means Clustering*

Selanjutnya dilakukan *clustering* data menggunakan *K-Means Clustering* yaitu dimulai dengan menentukan jumlah *cluster*, menentukan *centoroid* data secara acak, menghitung jarak dengan rumus *euclidean distance* hingga suata data tersebut tidak berubah atau berpindah *cluster*.

3.4.1 Menentukan Jumlah *Cluster*

Langkah awal yang dilakukan adalah menentukan jumlah *cluster* yang akan diuji. Pada penelitian ini dibagi menjadi 5 *cluster*. Penentuan *cluster* tersebut didasarkan atas pertimbangan kompleksitas dan kemudahan dalam interpretasi. Jika ada terlalu banyak kluster, hasil *clustering* bisa menjadi terlalu kompleks dan sulit diinterpretasi. Sementara itu, jika jumlah kluster terlalu sedikit, kluster bisa menjadi terlalu umum. Selain atas pertimbangan tersebut, pertimbangan lain adalah dengan mempertimbangan nilai *silhouette score*nya yaitu seberapa mirip setiap titik data dengan *clusternya* sendiri dibandingkan dengan kluster lain.

3.4.2 *Centroid Data*

Selanjutnya adalah menentukan nilai *centroid* awal yang didapatkan secara acak pada nilai variabel data dalam *cluster* sebanyak jumlah *cluster* yang telah ditentukan sebelumnya. Setelah ditentukan nilai *centroid* awal secara acak, dan melalui proses iterasi maka diperoleh nilai *centroid* akhir pada Tabel 4 yaitu sebagai berikut:

Tabel 4. *Centroid Data Akhir*

Jumlah <i>Cluster</i>	Jumlah Kasus	Jumlah Kematian	...	Luas Wilayah	Luas Wilayah Tegal/ Kebun
C1	3.75	0.00	...	1.83	2.62
C2	1.29	1.43	...	6.35	9.42
C3	4.62	3.08	...	1.58	4.03
C4	8.74	8.18	...	1.15	3.14
C5	1.67	5.55	...	2.37	4.22

Berdasarkan Tabel 4 didapatkan hasil bahwa untuk variabel jumlah kasus pada $C4 > C3 > C1 > C5 > C2$. Hal ini menunjukkan bahwa jumlah kasus sangat rawan berada di C4, rawan di C3, cukup aman di C1, aman di C5, dan sangat aman di C2, pada variabel jumlah kematian diketahui $C4 > C5 > C3 > C2 > C1$ yang memiliki arti bahwa jumlah kematian sangat rawan berada di C4, rawan di C5, cukup aman di C3, aman di C2, dan sangat aman di C1, selanjutnya pada variabel CFR diketahui bahwa $C4 > C2 > C5 > C3 > C1$ menandakan bahwa variabel CFR sangat rawan berada di C4, rawan di C2, cukup aman di C5, aman di C3, dan sangat aman di C1, pada variabel jumlah penduduk diketahui bahwa $C4 > C5 > C2 > C1 > C3$. Hal ini menunjukkan bahwa variabel jumlah penduduk sangat rawan berada di C4, rawan di C5, cukup aman di C2, aman di C1, dan sangat aman di C3, selanjutnya pada variabel jumlah petani yaitu diketahui $C1=9.95 > C4 = 8.66 > C2 = 4.24 > C5 = 1.23 > C3 = 1.14$ hal ini berarti bahwa jumlah petani sangat rawan berada di C1, rawan di C4, cukup aman di C2, aman di C5, dan sangat aman di C3, pada variabel jumlah banjir diketahui bahwa $C1 = 2.75 > C4 = 1.82 > C3 = 1.00 > C2 = 0.00$ dan $C5 = 0.00$ hal ini menunjukkan pada variabel jumlah banjir sangat rawan berada di C1, rawan di C4, cukup aman di C3, aman dan sangat aman berada di C2 dan C5,

selanjutnya pada variabel jumlah tanah longsor diketahui bahwa $C1 = 4.25 > C4 = 3.64 > C3 = 2.54 > C2 = 2.22 > C5 = 0.00$ hal ini menandakan bahwa jumlah tanah longsor sangat rawan berada di C1, rawan di C4, cukup aman di C3, aman di C2, dan sangat aman di C5, pada variabel luas wilayah diketahui bahwa $C2 = 6.35 > C5 = 2.37 > C1 = 1.83 > C3 = 1.58 > C4 = 1.15$. Hal ini menunjukkan bahwa luas wilayah sangat rawan berada di C2, rawan di C5, cukup aman di C1, aman di C3, dan sangat aman di C4, pada variabel luas wilayah kebun/tegal diketahui bahwa $C2 = 9.42 > C5 = 4.22 > C3 = 4.03 > C4 = 3.14 > C1 = 2.62$ yang berarti bahwa luas wilayah kebun/tegal sangat rawan berada di C2, rawan di C5, cukup aman di C3, aman di C4, sangat aman di C1.

Berdasarkan penjelasan tersebut dapat diketahui pola variabel terbanyak pada masing-masing *cluster* dan dapat disimpulkan bahwa pada C1 menghasilkan paling banyak kategori sangat aman, C2 paling banyak dengan kategori aman, C3 paling banyak dengan kategori cukup aman, C4 paling banyak dengan kategori sangat rawan, dan C5 paling banyak dengan kategori rawan.

3.4.3 Perhitungan Jarak dengan *Euclidean Distance*

Euclidean Distance adalah metode dengan tujuan mencari jarak dalam suatu kelompok. Dengan menggunakan perhitungan *Euclidean Distance* akan diperoleh hasil data Leptospirosis dengan jarak terdekat berdasarkan 5 nilai *centroid* yang berbeda. Data yang ada akan dibagi menjadi beberapa *cluster*, suatu *cluster* akan diperoleh dari jarak antar pusat *cluster* yang nantinya akan dijadikan sebagai nilai acuan suatu iterasi. Lalu pada iterasi selanjutnya, perhitungannya diperoleh dengan mengubah nilai *centroid* rata-rata. Selanjutnya data pada setiap *cluster* pada iterasi pertama dan kedua akan dibandingkan, apabila terdapat perbedaan *cluster* pada kedua iterasi tersebut, maka iterasi akan terus dilanjutkan dengan menentukan pusat *cluster* baru dan melakukan perhitungan data kembali. Hal ini terus dilakukan hingga suatu data tersebut tidak berubah atau berpindah *cluster*. Hasil clustering terdapat pada Tabel 5.

Tabel 5: Hasil *Clustering*

No.	Kabupaten/Kota	C1	C2	C3	C4	C5
1.	Pacitan				✓	
2.	Kab. Kediri	✓				
...
37.	Kota Batu		✓			
38.	Kota Surabaya					✓

Berdasarkan Tabel 5 didapatkan hasil *clustering* wilayah rawan Leptospirosis di Jawa Timur terdapat 4 wilayah yang termasuk kedalam C1 dengan kategori sangat aman yaitu Kab. Kediri, Kab. Banyuwangi, Kab. Pasuruan, dan Kab. Sidoarjo. Terdapat 7 wilayah termasuk C2 dengan kategori aman yaitu Kota Kediri, Kota Blitar, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, dan Kota Batu. Selanjutnya, terdapat 13 wilayah yang termasuk kedalam C3 dengan kategori cukup aman yang terdiri dari Kab. Tulungagung, Kab. Blitar, Kab. Lumajang, Kab. Probolinggo, Kab. Mojokerto, Kab. Jombang, Kab. Nganjuk, Kab. Bojonegoro, Kab. Tuban, Kab. Lamongan, Kab. Gresik, Kab. Bangkalan, dan Kab. Sumenep. Terdapat 11 wilayah termasuk kedalam C4 dengan kategori sangat rawan yaitu meliputi wilayah Kab. Pacitan, Kab. Ponorogo, Kab. Trenggalek, Kab. Bondowoso, Kab. Situbondo, Kab. Madiun, Kab. Magetan, Kab. Ngawi, Kab. Sampang, Kab. Pamekasan, dan Kota Malang. Selanjutnya terdapat 3 wilayah termasuk kedalam C5 dengan kategori rawan yaitu terdiri dari wilayah Kab. Malang, Kab. Jember, dan Kota Surabaya.

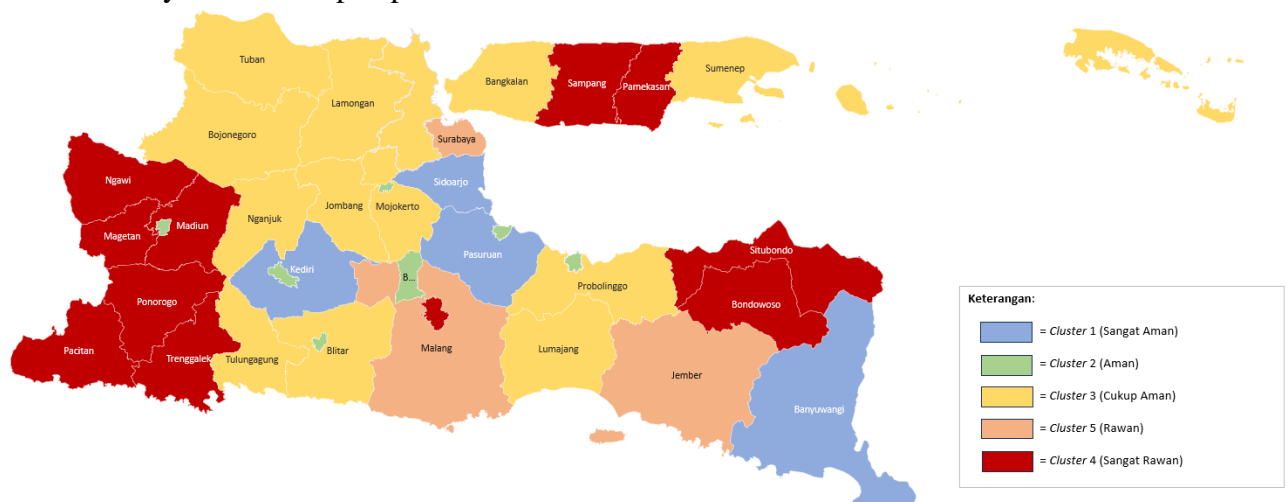
3.5 Evaluasi *Silhouette Coefficient*

Setelah itu, dilakukan tahap evaluasi atau uji akurasi data menggunakan *Silhouette Coefficient* yang bertujuan untuk mengetahui kualitas hasil *clustering* data atau mengetahui *cluster* yang paling optimal. Berikut hasil *Silhouette Coefficient* pada Tabel 6.

Tabel 6. Hasil *Silhouette Coefficient*

Jumlah Cluster	Hasil <i>Silhouette Coefficient</i>
2	0.49
3	0.56
4	0.61
5	0.65

Berdasarkan hasil pada Tabel 6. didapatkan bahwa *cluster* 5 memiliki nilai yang paling tinggi diantara lainnya. Pemilihan jumlah *cluster* sebanyak 5 pada penelitian ini sudah tepat, karena hasil dari evaluasi *Silhouette Coefficient* menghasilkan nilai yang paling baik diantara jumlah *cluster* lainnya yaitu sebesar 0.65. Nilai tersebut menyatakan bahwa hasil *clustering* yang telah dilakukan masuk kedalam kategori baik. Berikut ditampilkan Gambar 2 terkait pemetaan wilayah rawan Leptospirosis di Jawa Timur tahun 2023.



Gambar 2. Peta Cluster Wilayah Rawan Leptospirosis di Jawa Timur Tahun 2023

Sumber: Hasil olahan penulis

Penelitian ini menghasilkan suatu *clustering* wilayah yang terbagi menjadi 5 kategori cluster dengan kategori sangat aman, aman, cukup aman, rawan, dan sangat rawan, apabila dibandingkan dengan penelitian yang dilakukan oleh Nurvita dengan hasil akhir yaitu diketahui bahwa morbiditas dan mortalitas Leptospirosis yang tersebar di empat kelurahan wilayah Puskesmas Kedungmundu masih belum terbentuk suatu pemetaan wilayah [16]. Penelitian ini juga menghasilkan pemetaan wilayah rawan Leptospirosis di wilayah Jawa Timur menggunakan metode *K-Means Clustering* disertai dengan 9 variabel pendukung dan menghasilkan evaluasi *clustering* dengan kategori baik, dibandingkan dengan penelitian yang dilakukan oleh Damayanti terkait penentuan tingkat kerawanan penyebaran Leptospirosis menggunakan metode inferensi *Fuzzy Tsukamoto* yang masih terbatas pada wilayah di Kabupaten Bantul dan hanya menggunakan 5 variabel pendukung dan tidak terdapat hasil evaluasi *clustering* [18]. Hasil penelitian lainnya yang dilakukan oleh Rohman, dkk. terkait pengelolaan wisata air berwawasan kesehatan pemetaan kasus Leptospirosis menghasilkan sebanyak 2 cluster yaitu tinggi dan rendah [35], hasil *clustering* yang ditampilkan pada

penelitian ini hanya terbatas pada wilayah Kabupaten Klaten, tidak terdapat penjelasan terkait evaluasi *clustering*, dan tidak terbentuknya suatu peta hasil *clustering*.

4 Kesimpulan

Penelitian ini memperoleh hasil yaitu tingkat kerawanan Leptospirosis di Jawa Timur Tahun 2023 terbagi menjadi 5 *cluster* wilayah yaitu *cluster* sangat aman (C1) sebanyak 4 Kabupaten/Kota, *cluster* aman (C2) sebanyak 7 Kabupaten/Kota, *cluster* cukup aman (C3) sebanyak 13 Kabupaten/Kota, *cluster* rawan (C5) sebanyak 3 Kabupaten/Kota, dan *cluster* sangat rawan (C4) sebanyak 11 Kabupaten/Kota di Jawa Timur Tahun 2023 dan didapatkan nilai akurasi menggunakan *Silhouette Coefficient* sebesar 0.65 dengan kategori baik. Penelitian ini diharapkan dapat membantu pemerintah dan pihak terkait dalam mengatasi kasus Leptospirosis khususnya di Jawa Timur. Selain itu diharapkan juga dapat meningkatkan kesadaran dan kewaspadaan terhadap kasus Leptospirosis bagi masyarakat sekitar sehingga kasus Leptospirosis di tahun selanjutnya mengalami penurunan.

Daftar Pustaka

- [1] F. A. Sukma, M. Mulida, and K. S. Sujana, "Severe Leptospirosis," *Cermin Dunia Kedokt.*, vol. 48, no. 11, p. 353, 2021, doi: 10.55175/cdk.v48i11.1556.
- [2] H. S. Wulandari and L. Y. Hendrati, "Investigasi Kasus Kejadian Luar Biasa Leptospirosis Di Probolinggo Tahun 2022," *Care J. Ilm. Ilmu Kesehat.*, vol. 10, no. 3, pp. 390–400, 2022, doi: 10.33366/jc.v10i3.3739.
- [3] S. A. Sijid, C. Muthiadin, Z. Zulkarnain, and R. Adi Purba, "Faktor-Faktor Yang Berpengaruh Terhadap Kejadian Leptospirosis Dan Pencegahannya (Review)," *Teknosains Media Inf. Sains dan Teknol.*, vol. 16, no. 2, pp. 214–220, 2022, doi: 10.24252/teknosains.v16i2.28154.
- [4] M. Subuh, "Petunjuk Teknis Pengendalian Pes," p. 11, 2014, [Online]. Available: https://ptvz.kemkes.go.id/storage/media-download/file/file_1619051335.pdf
- [5] K. W. D. Nugraha, *Profil Kesehatan Indonesia 2022*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2022.
- [6] S. J. Jahja and C. Drew, "Peningkatan Pengetahuan dalam Upaya Pencegahan Kasus Baru Leptospirosis di Wilayah Kerja Puskesmas Kresek," *Malahayati Nurs. J.*, vol. 6, no. 2, pp. 725–735, 2024, doi: 10.33024/mnj.v6i2.12875.
- [7] F. D. Handayani *et al.*, "Penguatan Kebijakan One Health dan Jejaring Laboratorium Dalam Deteksi Dini Leptospirosis di Indonesia," *Bul. Penelit. Kesehat.*, vol. 47, no. 4, 2019, doi: 10.22435/bpk.v47i4.1928.
- [8] P. Alkhairi and A. P. Windarto, "Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara," *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.
- [9] R. D. Christyanti, A. Arif, A. P. Utomo, and M. Ayyub, "Implementasi Metode Fuzzy C-Means dalam Clustering Wilayah Rawan Penyakit Demam Berdarah," *J. Math. Educ. Sci.*, vol. 6, no. 1, pp. 11–17, 2022, doi: 10.32665/james.v6i1.933.
- [10] Fithriyah Azzahrah, S. Annas, and Z. Rais, "Hybrid Hierarchical Clustering dalam Pengelompokan Daerah Rawan Bencana Tanah Longsor di Sulawesi Selatan," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 153–161, 2022, doi: 10.35580/variansiunm38.
- [11] D. Setiadi *et al.*, "Penerapan Algoritma K-Means Clustering Pada Pengelompokan Daerah Penyebaran Diare di Provinsi Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3320–3327, 2023.
- [12] M. A. Sembiring, "Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (Dbd)," *J. Sci. Soc. Res.*, vol. 4, no. 3,

- p. 336, 2021, doi: 10.54314/jssr.v4i3.712.
- [13] R. Niken, E. Tiasti, A. Hamid, D. Candra, and R. Novitasari, "Implementasi Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Penyakit Diabetes Melitus Di Kecamatan Bagor," *J. Mhs. Mat. Algebr.*, vol. 02, no. 01, pp. 69–84, 2021.
- [14] A. Rizal, D. C. R. Novitasari, and M. Hafiyusholeh, "Pengelompokan Karyawan Berdasarkan Kesalahan Menggunakan Perbandingan Fuzzy C-Means, K-Means, dan Probabilistic Distance Clustering," *J. Fourier*, vol. 11, no. 2, pp. 69–77, 2022, doi: 10.14421/fourier.2022.112.69-77.
- [15] A. L. R. Putri and N. Dwidayati, "Analisa Perbandingan K-Means dan Fuzzy C-Means dalam Pengelompokan Daerah Penyebaran Covid-19 Indonesia," *UNNES J. Math.*, vol. 10, no. 2, pp. 50–55, 2021, [Online]. Available: <http://journal.unnes.ac.id/sju/index.php/ujm>
- [16] S. Nurvita, "Pemetaan Epidemiologi Leptospirosis Berbasis Sistem Informasi Geografis Tahun 2021-2022," *Prepotif J. Kesehat. Masy.*, vol. 7, no. 3, pp. 16754–16761, 2023, [Online]. Available: <http://journal.universitaspahlawan.ac.id/index.php/prepotif/article/view/21865>
- [17] I. T. Ganinov and S. Huda, "Penerapan Sistem Informasi Geografis Faktor Risiko Penyakit Leptospirosis," *J. Ilm. Ilmu Kesehat. Wawasan Kesehat.*, vol. 5, no. 2, pp. 280–284, 2019, doi: 10.33485/jiik-wk.v5i2.143.
- [18] A. Damayanti, "Penentuan Tingkat Kerawanan Penyebaran Leptospirosis Menggunakan Inferensi Fuzzy Tsukamoto," *J. Sist. Komput. dan Inform.*, vol. 1, no. 1, p. 52, 2019, doi: 10.30865/json.v1i1.1388.
- [19] F. Cilia, Giovanni; Bertelloni, S. Albini, and F. Fratini, "Insight into the Epidemiology of Leptospirosis : A Review of," *Animals*, vol. 11, no. 191, pp. 1–16, 2021.
- [20] I. Tolistiawaty, N. Hidayah, and A. N. Widayati, "Faktor Lingkungan Abiotik dan Kejadian Leptospirosis Pada Tikus di Desa Lalombi Kabupaten Donggala Sulawesi Tengah," *Pemakalah Paralel*, no. November 2019, pp. 119–123, 2020.
- [21] A. V. Samrot *et al.*, "Leptospiral infection, pathogenesis and its diagnosis—a review," *Pathogens*, vol. 10, no. 2, pp. 1–30, 2021, doi: 10.3390/pathogens10020145.
- [22] I. Ningsih and M. H. Wahid, "Leptospirosis Ditinjau dari Aspek Mikrobiologi," *EKOTONIA J. Penelit. Biol. Bot. Zool. dan Mikrobiol.*, vol. 7, no. 1, pp. 31–43, 2022, doi: 10.33019/ekotonia.v7i1.3141.
- [23] S. Eka Purnama, B. Hartono, P. Studi Magister Ilmu Kesehatan Masyarakat, and F. Kesehatan Masyarakat, "Faktor Risiko Kejadian Leptospirosis Di Indonesia: Literature Review," *J. Kesehat. Masy.*, vol. 6, no. 3, pp. 2010–2022, 2022.
- [24] A. Nugroho, M. S. Adi, and Nurjazuli, "Jurnal Ilmu Kesehatan Masyarakat Penularan Leptospirosis di Indonesia : Literature Review," *J. Ilmu Kesehat. Masy.*, vol. 12, no. 123, pp. 57–64, 2023.
- [25] C. Goarant, G. Trueba, E. Bierque, R. Thibeaux, and A. De Pena-moctezuma, "Leptospira and Leptospirosis," *Water Sanit. 21stCentury Heal. Microbiol. Asp. Excreta Wastewater Manag. (Global WaterPathogen Proj.*, no. June 2016, 2019.
- [26] A. D. Indriyanti, D. R. Prehanto, and T. Z. Vitadiar, "K-means method for clustering learning classes," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 22, no. 2, pp. 835–841, 2021, doi: 10.11591/ijeecs.v22.i2.pp835-841.
- [27] N. K. Zuhal, "Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering," *Pros. Semin. Nas. Teknol. dan Sains*, vol. 1, pp. 200–205, 2022, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/stains/article/view/1495>
- [28] N. Ulinnuha and S. A. Sholihah, "Analisis Cluster untuk Pemetaan Data Kasus Covid-19 di Indonesia Menggunakan K-Means," *J. Mat. dan Stat. serta Apl.*, vol. 9, no. 2, 2021.
- [29] A. Fira, C. Rozikin, and G. Garno, "Komparasi Algoritma K-Means dan K-Medoids

- Untuk Pengelompokan Penyebaran Covid-19 di Indonesia,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 133–138, 2021, doi: 10.30871/jaic.v5i2.3286.
- [30] N. Ulinnuha, “Provincial Clustering in Indonesia Based on Plantation Production Using Fuzzy C-Means,” *ITSMART J. Ilm. Teknol. dan Inf.*, vol. 9, no. 1, pp. 8–12, 2020.
- [31] M. Shutaywi and N. N. Kachouie, “Silhouette analysis for performance evaluation in machine learning with applications to clustering,” *Entropy*, vol. 23, no. 6, pp. 1–17, 2021, doi: 10.3390/e23060759.
- [32] A. Atira and B. Nurina Sari, “Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Cluster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan,” *J. Ilm. Wahana Pendidik.*, vol. 9, no. 17, pp. 76–86, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.8282638>