

Optimasi Algoritma K-Means Menggunakan Metode Elbow Untuk Pengelompokan Penyakit Menular Di Puskesmas Cigugur Tengah

Ayu Hendrati Rahayu¹, Sabila Nopiyanti²

^{1,2}Program Studi Teknik Informatika, Politeknik TEDC Bandung

^{1,2}Jln. Pesantren-Cibabat, Kota Cimahi, 40513, Indonesia

E-mail: ayuhendrati@poltektedc.ac.id¹, sabilanpy03@gmail.com²

Abstrak

Info Naskah:

Naskah masuk:

11 Desember 2024

Direvisi:

7 Maret 2025

Diterima:

7 Maret 2025

Penyakit menular masih menjadi masalah utama di Indonesia, terutama di wilayah Puskesmas Cigugur Tengah, Kota Cimahi, dengan berbagai penyakit seperti Tuberkulosis, ISPA, dan demam berdarah yang dilaporkan. Namun, pengelolaan data penyakit menular masih terbatas pada analisis sederhana, sehingga diperlukan metode yang lebih efektif untuk meningkatkan ketepatan dalam pengambilan kebijakan penanggulangan penyakit. Penelitian ini menerapkan Algoritma K-Means dengan metode *elbow* untuk mengelompokkan data penyakit menular guna mendapatkan kluster optimal. Berdasarkan hasil pengujian, diperoleh tiga kluster optimal ($k=3$). *Cluster_0* mencakup pasien berusia 17-25 tahun dengan dominasi penyakit *Acute Nasopharyngitis* (J00) di wilayah RT 10, RW 13, dan 14. *Cluster_1* memiliki karakteristik serupa dengan dominasi penyakit *Acute Upper Respiratory* (J06.9) di RT 1, RW 3. *Cluster_2* berfokus pada anak-anak usia 0-5 tahun dengan dominasi penyakit *Acute Nasopharyngitis* (J00) di RT 1, RW 14. Analisis ini memberikan jarak rata-rata ke *centroid* sebesar 6038,938 dan nilai *Davies-Bouldin Index* 0,507, menunjukkan kualitas klusterisasi yang baik. Hasil klusterisasi ini diharapkan membantu Puskesmas Cigugur Tengah dalam merancang strategi intervensi yang lebih tepat sasaran dengan memanfaatkan data secara efektif, mengingat keterbatasan sumber daya dan variabilitas pola penyebaran penyakit di daerah tersebut.

Abstract

Keywords:

Infectious Diseases;

K-Means;

Elbow Method;

Clustering;

Cigugur Tengah Health Center.

Infectious diseases remain a significant public health challenge in Indonesia, especially in the Cigugur Tengah Health Center area, Cimahi City, where various diseases like Tuberculosis, ARI, and dengue fever are reported. However, infectious disease data management is still limited to simple analysis, necessitating a more effective method to enhance decision-making accuracy in disease prevention strategies. This study applies the K-Means algorithm with the elbow method to cluster infectious disease data to determine the optimal clusters. Testing results indicate an optimal cluster count of three ($k=3$). *Cluster_0* includes patients aged 17-25 years with a dominant disease, *Acute Nasopharyngitis* (J00), in RT 10, RW 13, and 14 areas. *Cluster_1* shows similar demographics with a dominant disease, *Acute Upper Respiratory* (J06.9), in RT 1, RW 3. *Cluster_2* focuses on children aged 0-5 years with a dominant disease, *Acute Nasopharyngitis* (J00), in RT 1, RW 14. This analysis provides an average centroid distance of 6038.938 and a *Davies-Bouldin Index* of 0.507, indicating good clustering quality. The clustering results are expected to assist Cigugur Tengah Health Center in designing more targeted intervention strategies by leveraging data effectively, given the resource limitations and varying patterns of disease spread in the region.

*Penulis korespondensi:

Sabila Nopiyanti

E-mail: sabilanpy03@gmail.com

1. Pendahuluan

Penyakit menular masih menjadi masalah utama kesehatan masyarakat di Indonesia dan masih sering timbul sebagai Kejadian Luar Biasa (KLB) yang menyebabkan kematian bagi penderitanya. Penyakit menular adalah penyakit yang dapat ditularkan secara langsung atau melalui perantara dari orang yang sakit ke orang yang sehat atau orang yang tidak tertular penyakit menular. Di Kota Cimahi, penyakit menular juga menjadi perhatian serius. Berdasarkan laporan Dinas Kesehatan Kota Cimahi pada tahun 2022, tercatat beberapa jenis penyakit yaitu seperti *Tuberkulosis* dengan kasus 4.603, ISPA 2.800 kasus dan demam berdarah sekitar 180 kasus yang dilaporkan dari Puskesmas setempat [1].

Berdasarkan hasil observasi awal yang dilakukan selama 1 (satu) minggu didapatkan bahwa di Puskesmas Cigugur Tengah, terdapat 803 penyakit menular dari 15 jenis penyakit menular yang ada di wilayah tersebut. Namun seiring dengan meningkatnya frekuensi kasus serta pengolahan data yang masih menggunakan analisis sederhana tentunya diperlukan penanganan khusus mengenai pengolahan data penyakit menular, sehingga kebijakan yang diambil terkait langkah-langkah penanggulangan penyakit menular dapat lebih efektif dan tepat sasaran. Salah satu pendekatan yang efektif dalam menganalisis data tersebut adalah melalui data mining.

Data *mining* memiliki beberapa teknik yang terkenal dan sering digunakan oleh peneliti, diantaranya seperti *clustering*, *classification*, *association*, dan beberapa perkembangan teknik sesuai dengan perubahan kecenderungan data pada saat ini [2]. Pada kali ini peneliti menggunakan teknik *Clustering* untuk mengelompokkan penyakit menular pada Puskesmas Cigugur Tengah.

Dalam penelitian ini juga peneliti menggunakan algoritma *K-means*, yang mana *K-Means Clustering* merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil. Algoritma ini mampu meminimalkan jarak antara data ke *clusternya* [3][4].

Algoritma *K-Means* sendiri pernah di gunakan oleh beberapa peneliti salah satunya adalah Sugianto menggunakan algoritma *K-Means* untuk mengelompokkan data penyakit dengan hasil nilai *Davies Bouldin Index* untuk algoritma *K-Means* sebesar -0.453 dan algoritma *K-Medoids* sebesar -1.276. Maka dapat disimpulkan bahwa algoritma *k-means* lebih baik dari algoritma *K-Medoids* yang menghasilkan nilai rata – rata *Davies Bouldin Index* sebesar -1.276 [5]. Dan dilakukan oleh Olivia yang menghasilkan lima *cluster* daerah penyakit yang banyak diderita oleh pasien di Daerah penelitiannya. Hasil pengujian pada sistem tersebut mempunyai hasil nilai SSE, diperoleh nilai *error* terendah pada *cluster* 3 yaitu sebesar $1,68497 \times 10^{14}$ dengan mendapatkan 3 *cluster* daerah terbaik [6].

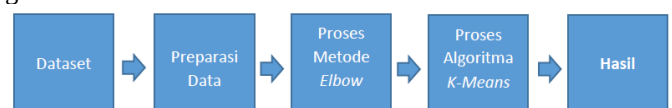
Namun, penentuan jumlah *cluster* yang optimal seringkali menjadi tantangan, karena jumlah *cluster* yang dipilih secara manual bisa menyebabkan hasil pengelompokan yang kurang akurat, maka di digunakanlah metode *elbow* sebagai metode untuk menentukan K terbaik.

Metode *elbow* merupakan salah satu metode yang dapat digunakan untuk menentukan jumlah *cluster* terbaik, yaitu dengan cara melihat presentase setiap *cluster* yang akan membentuk siku pada suatu titik tertentu. Metode *elbow* biasa disajikan dalam bentuk grafik untuk mengetahui lebih jelas siku yang terbentuk. Tujuan dari metode *elbow* adalah untuk memilih nilai k yang kecil dan masih memiliki nilai *withinss* yang rendah [7]. Metode *elbow* sendiri pernah digunakan oleh beberapa peneliti salah satunya adalah [8] menggunakan optimisasi algoritma *K-Means* menggunakan metode *elbow* untuk pengelompokan data stunting dengan hasil *cluster* terbaik $K=4$ dengan nilai 12.869 dan hasil *cluster* pada data ini yaitu 4 kelompok dengan *Cluster_0* dengan jumlah 98, *Cluster_1* dengan jumlah 89, *Cluster_2* dengan jumlah 70 dan *Cluster_3* dengan jumlah 63 data dengan total 320 data dan nilai *avg. within centroid distance* adalah 12.869 sedangkan nilai *Davies Bouldin Index* 0.150.

Penggunaan algoritma *K-Means* dengan metode *elbow* untuk mengelompokkan penyakit menular dapat digunakan untuk mengoptimalkan hasil klasterisasi dan membantu dalam mengidentifikasi pola penyebaran dan merancang strategi intervensi yang lebih efektif. Hal ini bertujuan untuk memberikan informasi yang lebih baik dalam pengambilan keputusan di Puskesmas yang menghadapi beberapa tantangan, antara lain yaitu volume Data yang Besar sehingga Puskesmas memerlukan proses dan analisis data secara efektif, Keterbatasan dalam sumber daya manusia dan fasilitas menyulitkan analisis dan pengelompokan data penyakit menular secara manual. Serta perbedaan Pola Penyakit yang mana Penyebaran penyakit menular dapat bervariasi secara signifikan antara daerah, sehingga memerlukan pendekatan yang lebih terarah dan berbasis data. Berdasarkan uraian latar belakang di atas, peneliti tertarik mengangkat topik penelitian yang berjudul “Optimasi Algoritma *K-Means* Menggunakan Metode *Elbow* untuk Pengelompokan Penyakit Menular di Puskesmas Cigugur Tengah”.

2. Metode

Metode yang digunakan dalam pengolahan data dengan algoritma *K-Means* pada penelitian ini, dapat dilihat gambar 1.



Gambar 1. Contoh Penggunaan

Berdasarkan gambar 1, terdapat beberapa tahapan yaitu diantaranya sebagai berikut.

2.1 Dataset

Dataset yang digunakan dalam penelitian ini adalah data penyakit yang diambil langsung dari Puskesmas Cigugur Tengah dengan persetujuan dari Dinas Kesehatan

Kota Cimahi. Data yang diambil merupakan data penyakit menular yang berjumlah (803 data) dari 15 jenis penyakit menular dari periode Juli-Agustus 2024. Data penyakit menular yang ada di Puskesmas Cigugur Tengah bisa dilihat pada tabel 1 dibawah ini.

Tabel 1. Data Penyakit Menular Di Puskesmas Cigugur Tengah

Nama Penyakit	Kode Penyakit
Diare dan Gastroenteritis	A09
DHF	A91
Respiratory Tuberculosis	A15
Varicella	B01
Herpes Zoster	B02
HIV	B20
Mumps	B26
Dermatophytosis	B35
Scabies	B86
Impetigo	I01
Acute Nasopharyngitis	J00
Acute Pharyngitis	J02
Acute Upper Respiratory	J06.9
Bronchopneumonia	J18.0
Pneumonia	J18.9

2.2 Preparasi Data

Data yang sudah didapatkan selanjutnya akan dipreparasi atau cleaning data. Dimana proses ini, data akan diolah kembali untuk menghilangkan data yang tidak konsisten atau menghapus data yang kurang penting pada proses pengolahan data nantinya. Sehingga nanti akan menghasilkan data yang benar-benar siap untuk di *clustering* pada proses selanjutnya. *Field* yang digunakan dari data tersebut adalah *field* umur pasien, jenis kelamin, kode penyakit, kelurahan, RT, RW dan jenis kunjungan. Dimana *field-field* tersebut akan penulis gunakan sebagai indikator analisis untuk hasil penelitian. Sampel data yang telah dipreparasi dapat dilihat pada tabel 2.

Tabel 2. Sampel Data Yang Sudah Di Preparasi

Id	Umur	Jenis Kelamin	Kode Penyakit	Kelurahan	RT	RW	Jenis Kunjungan
1	3	0	0	3	4	15	1
2	1	0	1	3	3	19	0
3	4	0	2	3	2	11	1
4	2	1	3	3	4	2	0
5	4	1	4	3	5	1	1
6	4	0	5	3	7	3	0
7	5	0	6	3	7	5	0
8	1	0	7	3	2	3	1
9	3	1	8	3	2	5	0
10	4	1	9	3	5	7	0
11	5	0	10	3	5	2	0
12	3	0	11	3	4	4	0
13	4	1	12	3	4	5	0
14	7	0	13	3	5	15	0
15	5	1	14	3	2	1	1

2.3 Proses Metode Elbow

Metode *elbow* merupakan sebuah teknik untuk menentukan *cluster* yang paling optimal dengan cara memperhatikan perbandingan hasil antara jumlah *cluster* yang membentuk siku pada suatu titik. Metode ini melibatkan dua langkah utama [9]:

1) Perhitungan Inertia

Inertia adalah jumlah kuadrat jarak antara data dan pusat kluster. Inertia yang lebih rendah menunjukkan kluster yang lebih baik. Dalam penelitian ini, nilai inertia dihitung untuk setiap jumlah kluster (K) yang berbeda.

2) Plot Grafik Elbow

Setelah nilai inertia dihitung, dibuatlah grafik dengan sumbu X sebagai jumlah kluster (K) dan sumbu Y sebagai nilai inertia. Grafik ini digunakan untuk mengidentifikasi titik siku, di mana penurunan inertia mulai melambat. Titik ini menunjukkan jumlah kluster yang optimal. Hasil dari perhitungan masing-masing *cluster* dikelola di *Microsoft Excel* dengan memplot data ke dalam diagram kurva. Jika terdapat nilai yang paling banyak menyiku di antara hasil *cluster*, maka nilai tersebut dianggap sebagai jumlah *cluster* yang terbaik.

2.4 Proses Algoritma K-Means

Algoritma *K-Means* terdiri dari beberapa langkah yaitu: [10]:

1) Penentuan Jumlah Kluster

Jumlah kluster (K) yang optimal ditentukan terlebih dahulu menggunakan metode *elbow*. Setelah itu, *K-means* digunakan untuk mengelompokkan data penyakit menular berdasarkan kluster tersebut.

2) Penugasan Kluster

Dalam tahap ini, proses pengelompokan dilakukan dengan membangkitkan nilai random untuk pusat kluster awal (*centroid*) sebanyak K. Setiap data input kemudian diklasifikasikan berdasarkan kedekatannya dengan *centroid* menggunakan jarak *Euclidean* (*Euclidean Distance*). Rumus untuk jarak *Euclidean* adalah sebagai berikut:

$$(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \quad (1)$$

Keterangan:

x_i : Data kriteria

μ_j : *Centroid* pada kluster ke-j

3) Analisis Kluster

Setelah kluster terbentuk, setiap kluster dianalisis berdasarkan karakteristik, Usia, Jenis kelamin, kode ICD, Kelurahan, RT/RW, dan Jenis Kunjungan. Analisis ini membantu dalam memahami pola penyakit menular dan kelompok yang memerlukan perhatian khusus.

4) Visualisasi dan Interpretasi

Hasil pengelompokan kemudian divisualisasikan menggunakan teknik seperti *bar (column)* yang memudahkan dalam interpretasi. Penafsiran lebih lanjut dilakukan untuk mengidentifikasi area atau kelompok yang membutuhkan intervensi khusus.

5) Perbaikan Centroid

Nilai *centroid* akan diperbarui dengan mengambil rata-rata dari data dalam kluster yang bersangkutan, hingga pada iterasi terakhir nilai pusat kluster yang baru akan digunakan untuk klasifikasi *final*.

2.5 Hasil

Hasil yang diperoleh dalam penelitian ini mencakup pengelompokan data penyakit menular berdasarkan Usia, Jenis kelamin, kode ICD, Kelurahan, RT dan RW, dan Jenis Kunjungan di Kelurahan Cigugur Tengah. Pengelompokan dilakukan menggunakan algoritma *K- Means*, dan hasil penelitian ini akan mencakup evaluasi performa algoritma serta jumlah kluster terbaik yang diidentifikasi melalui metode *elbow*. Dari hasil maka akan memperoleh umpan balik dari praktisi, sebagai solusi atau upaya untuk mengatasi masalah yang ditimbulkan dari setiap kluster. Sebagai bagian dari validasi eksternal, peneliti akan mendapatkan umpan balik dari tenaga medis di Puskesmas Cimahi terkait hasil kluster yang dihasilkan.

3. Hasil dan Pembahasan

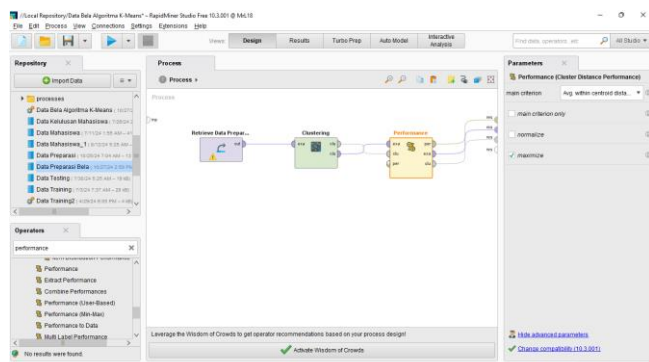
Berikut hasil dan pembahasan dari penelitian ini:

3.1 Hasil Penelitian

Hasil penelitian merupakan sub bagian yang membahas tiga tahapan yaitu menentukan jumlah *cluster* terbaik, pengujian menggunakan algoritma *k-means* menggunakan metode *elbow*, dan hasil data *cluster k-means*.

1) Menentukan Jumlah Cluster Terbaik

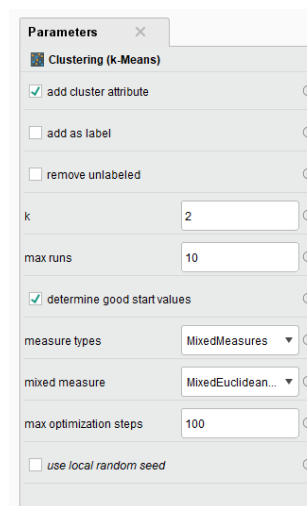
Tahap ini menggunakan metode *elbow* untuk menguji dan melihat hasil *cluster* terbaik dengan memasukkan data yang sudah dipreparasi sebelumnya dengan cara *import* data pada *software RapidMiner* dan menyeleksi data yang akan digunakan. Data yang akan diproses menggunakan *RapidMiner* dapat dilihat pada gambar 2 berikut.



Gambar 2. Pengolahan Data pada *RapidMiner*

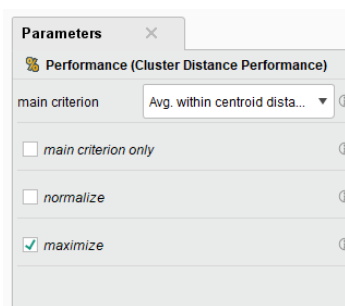
Berdasarkan gambar 2, data yang akan diuji akan dihubungkan dengan opsi *K-Means (Clustering)* lalu dihubungkan ke *Cluster Distance Performance* dan *Cluster Distance Performance* disambungkan ke *result*.

Selanjutnya mengatur dan menyesuaikan parameter *clustering* dengan jumlah pengujian yang akan dilakukan. Penulis akan melakukan 10 pengujian, maka nilai *k* diubah menjadi 2 hingga 11 untuk mendapatkan nilai *avg. within centroid distance* sebagai bahan untuk pengujian metode *elbow*. Dapat dilihat pada gambar 3 berikut.



Gambar 3. Mengatur Parameter *Clustering*

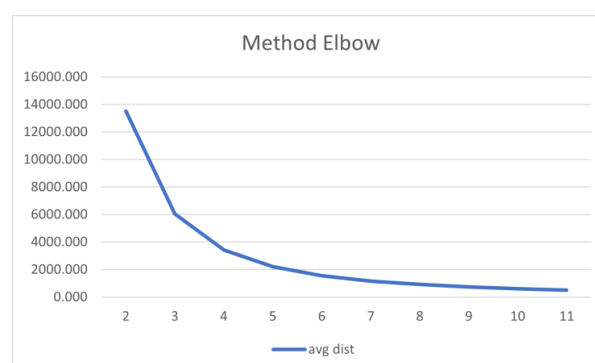
Lalu mengatur parameter dari *cluster distance performance*, dengan mencentang atau memilih kategori *maximize* agar dapat memperoleh *cluster distance performance* yang baik dan bernilai positif. Dapat dilihat pada gambar 4 berikut.



Gambar 4. Mengatur Parameter *Cluster Distance Performance*

2) Pengujian Algoritma *K-Means* Menggunakan Metode *Elbow*

Berdasarkan pengujian dengan menggunakan parameter *k=2* hingga *k=11*, maka nilai *avg. within centroid distance* akan dikonversi menjadi diagram untuk melihat nilai manakah yang memiliki kurva paling siku yang dihasilkan metode *elbow* untuk menentukan nilai *k* yang tepat pada *cluster* data. Berdasarkan pengujian dengan metode *elbow* maka ditemukan bahwa *k* terbaik adalah 3 dengan nilai *avg. within centroid distance* adalah 6038.938. Untuk hasilnya dapat dilihat pada gambar 5 berikut.



Gambar 5. Hasil Pengujian Metode *Elbow*

Data hasil pengujian menggunakan metode *elbow* dari 10 parameter *cluster* yang diuji dapat dilihat pada tabel 3 berikut.

Tabel 3. Hasil Pengujian Metode *Elbow*

No.	Jumlah Cluster	Nilai Avg. Within Centroid Distance	Davies-Bouldin Index (DBI)
1.	2	13501,219	0,504
2.	3	6038,938	0,507
3.	4	3426,443	0,513
4.	5	2217,871	0,518
5.	6	1561,115	0,524
6.	7	1164,649	0,532
7.	8	907,256	0,539
8.	9	731,531	0,547
9.	10	605,252	0,557
10	11	512,424	0,566

3) Hasil Data Cluster K-Means

Dari proses penentuan *cluster* didapatkan *cluster* 3 dengan menggunakan metode *elbow* dengan nilai yang terbaik. Maka pengelompokan data pada penelitian ini menjadi 3 kelompok atau *cluster* yaitu *cluster_0* dengan jumlah 268 data, kelompok data *cluster_1* berjumlah 267 data, dan kelompok data *cluster_2* berjumlah 268 data, dengan total 803 data dan nilai avg. within distance centroid adalah 6038,938 dan nilai Davies Bouldin Index adalah 0,507. Pengelompokan data tersebut dapat dilihat pada Gambar 6 dan Gambar 7 berikut.

Cluster Model

```
Cluster 0: 268 items
Cluster 1: 267 items
Cluster 2: 268 items
Total number of items: 803
```

Gambar 6. Hasil Pengujian Data Berdasarkan Cluster

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 6038.938
Avg. within centroid distance_cluster_0: 6061.307
Avg. within centroid distance_cluster_1: 5996.629
Avg. within centroid distance_cluster_2: 6058.720
Davies Bouldin: 0.507
```

Gambar 7. Hasil Performance Vector

3.2 Pembahasan

Pada bagian ini akan dibahas menjadi tiga tahapan yaitu analisis hasil algoritma *K-Means*, pengelompokan data penyakit menular dengan algoritma *K-Means*, dan upaya untuk mengurangi penyebaran penyakit menular.

1) Analisis Hasil Algoritma K-Means

Setiap atribut yang digunakan pada pengujian yang terdiri atas id, umur, jenis kelamin, kode penyakit, kelurahan, rt, rw dan jenis kunjungan, diperoleh data memperoleh nilai *maximum* dan *averagenya* yang dapat dilihat pada tabel 4 berikut.

Tabel 4. Analisis Nilai *Maximum* Dan *Average*

Atribut	Nilai Maximum	Nilai Average
Id	803	402

Atribut	Nilai Maximum	Nilai Average
Umur	8	3,503
Jenis Kelamin	1	0,491
Kode Penyakit	14	9,123
Kelurahan	24	3,158
RT	11	4,646
RW	23	10,821
Jenis Kunjungan	1	0,676

2) Pengelompokkan Data Penyakit Menular Menggunakan Algoritma K-Means

Dari hasil pengujian menggunakan algoritma *K-Means* didapatkan hasil *cluster* optimal yaitu $k=3$. Dengan demikian untuk pengelompokan *dataset* yang digunakan menjadi 3 kelompok *cluster*, dimana *cluster_0* sebanyak 268 data, *cluster_1* sebanyak 267 data, dan *cluster_2* sebanyak 268 data.

Untuk atribut umur pada *cluster_0* didominasi oleh umur 17-25 tahun (Remaja Akhir) dengan jumlah 51 data, untuk *cluster_1* didominasi oleh umur 17-25 tahun (Remaja Akhir) dengan jumlah 44 data, dan untuk *cluster_2* didominasi oleh umur 0-5 tahun (Balita) dengan jumlah 42 data. Untuk lebih jelasnya dapat dilihat pada tabel 5 berikut.

Tabel 5. Dominasi Atribut Umur

Umur	Kategori	Jumlah		
		Cluster_0	Cluster_1	Cluster_2
0-5	Balita	37	43	42
6-11	Kanak-Kanak	35	30	37
12-16	Remaja Awal	28	11	21
17-25	Remaja Akhir	51	44	37
26-35	Dewasa Awal	37	33	34
36-45	Dewasa Akhir	27	29	25
46-55	Lansia Awal	22	32	35
56-65	Lansia Akhir	18	25	21
>65	Manula	13	20	16

Untuk atribut jenis kelamin pada *cluster_0* didominasi oleh laki-laki dengan jumlah 137 data, untuk *cluster_1* didominasi oleh laki-laki dengan jumlah 142 data, dan untuk *cluster_2* didominasi perempuan dengan jumlah 138 data. Untuk lebih jelasnya dapat dilihat pada tabel 6 berikut.

Tabel 6. Dominasi Atribut Jenis Kelamin

Jenis Kelamin	Jumlah		
	Cluster_0	Cluster_1	Cluster_2
Laki-Laki	137	142	130
Perempuan	131	125	138

Untuk atribut kode penyakit pada *cluster_0* didominasi oleh penyakit *Acute Nasopharyngitis* (J00) dengan jumlah 137 data, untuk *cluster_1* didominasi oleh penyakit *Acute Upper Respiratory* (J06.9) dengan jumlah 96 data, dan untuk *cluster_2* didominasi oleh penyakit *Acute Nasopharyngitis* (J00) dengan jumlah 119 data. Untuk lebih jelasnya dapat dilihat pada tabel 7 berikut.

Tabel 7. Dominasi Atribut Kode Penyakit

Nama Penyakit	Kode Penyakit	Jumlah		
		Cluster_0	Cluster_1	Cluster_2
Diare dan Gastroenteritis	A09	20	0	10
DHF	A91	0	1	0
Respiratory Tuberculosis	A15	16	22	17
Varicella	B01	9	5	9
Herpes Zoster	B02	6	0	4
HIV	B20	3	1	1
Mumps	B26	15	10	4
Dermatophytosis	B35	1	11	4
Scabies	B86	5	0	2
Impetigo	I01	2	6	3
Acute Nasopharyngitis	J00	137	79	119
Acute Pharyngitis	J02	38	29	28
Acute Upper Respiratory	J06.9	13	96	65
Bronchopneumonia	J18.0	3	6	2
Pneumonia	J18.9	0	1	0

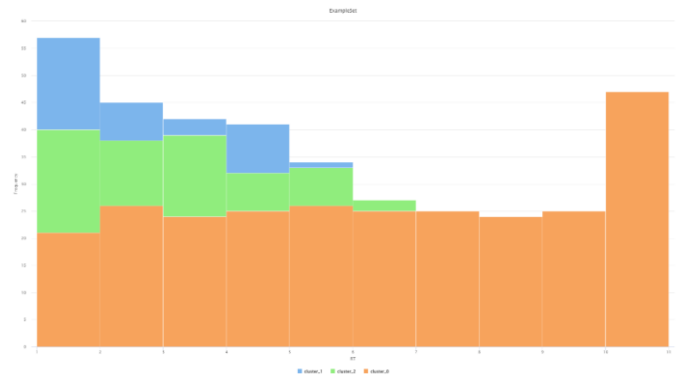
Untuk atribut keluhan pada *cluster_0* didominasi oleh keluhan Cigugur Tengah dengan jumlah 240 data, untuk *cluster_1* didominasi oleh keluhan Cigugur Tengah dengan jumlah 229 data, dan untuk *cluster_2* didominasi oleh keluhan Cigugur Tengah dengan jumlah 238 data. Untuk lebih jelasnya dapat dilihat pada tabel 8 berikut.

Tabel 8. Dominasi Atribut Keluhan

Keluhan	Jumlah		
	Cluster_0	Cluster_1	Cluster_2
Baros	14	31	12
Cibabat	3	0	1
Cibeber	3	1	1
Cigugur Tengah	240	229	238
Cimah	1	0	0
Karangmekar	1	2	0
Melong	0	0	1
Sukapura	0	0	1
Padasuka	1	1	3
Utama	0	1	0
Sukanampa	0	1	0
Cilame	0	1	0
Sindangkerta	0	0	1
Sariwangi	0	0	2
Gantar	0	0	1
Cipageran	0	0	1
Karanganyar	0	0	1
Lembang	0	0	1
Sarijadi	0	0	1
Cibeureum	0	0	1
Cangkorah	0	0	1
Rancakalong	0	0	1
Karang Mekar	0	1	0
Sayati	1	0	0
Petojo Selatan	1	0	0

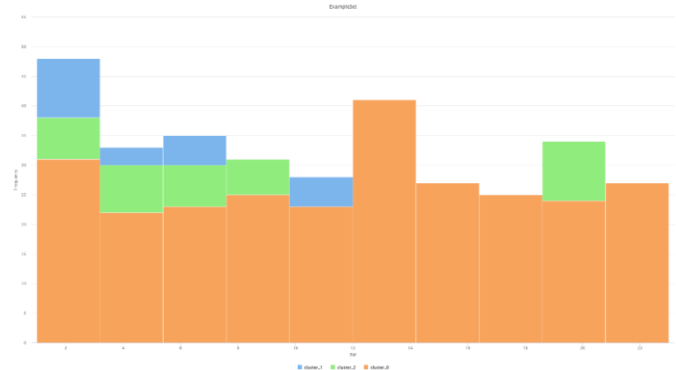
Untuk atribut RT pada *cluster_0* didominasi oleh RT 10 dengan jumlah 24 data, untuk *cluster_1* didominasi oleh RT 1

dengan jumlah 57 data, dan untuk *cluster_2* didominasi oleh RT 1 dengan jumlah 40 data. Untuk lebih jelasnya gambar 8 berikut.



Gambar 8. Dominasi Atribut RT

Untuk atribut RW pada *cluster_0* didominasi oleh RW 13 dan 14 dengan jumlah 14 data masing-masing, untuk *cluster_1* didominasi oleh RW 3 dengan jumlah 17 data, dan untuk *cluster_2* didominasi oleh RW 14 dengan jumlah 23 data. Untuk lebih jelasnya gambar 9 berikut.



Gambar 9. Dominasi Atribut RW

Untuk atribut jenis kunjungan pada *cluster_0* didominasi oleh pengunjung atau pasien baru dengan jumlah 144 data, untuk *cluster_1* didominasi oleh pengunjung atau pasien baru dengan jumlah 200 data, dan untuk *cluster_2* didominasi oleh pengunjung atau pasien dengan jumlah 199 data. Untuk lebih jelasnya dapat dilihat pada tabel 9 berikut.

Tabel 8. Dominasi Atribut Jenis Kunjungan

Jenis Kunjungan	Jumlah		
	Cluster_0	Cluster_1	Cluster_2
Lama	124	64	69
Baru	144	200	199

3) Upaya Mengurangi Penyebaran Penyakit Menular

Berdasarkan hasil analisis dan pengujian mengenai pengelompokan data penyakit menular menggunakan *K-Means clustering* diatas maka peneliti dapat menyimpulkan beberapa saran untuk menekan angka penyebaran penyakit menular di wilayah Puskesmas Cigugur Tengah.

- Edukasi kesehatan harus disesuaikan dengan usia dan penyakit dominan. Pada remaja (17-25 tahun) yang dominan di *cluster_0* dan *cluster_1*, fokuskan pada kebersihan diri, etika batuk, dan penggunaan masker untuk mencegah penyakit pernapasan. Media sosial

dan brosur di sekolah dapat menjadi sarana efektif. Sementara pada balita (0-5 tahun) di *cluster_2*, edukasi orang tua tentang kebersihan rumah, nutrisi yang cukup, dan menghindari kontak dengan orang sakit dapat mengurangi penyebaran penyakit.

- Pada laki-laki di *cluster_0* dan *cluster_1*, perlu meningkatkan kesadaran tentang hidup sehat, seperti menghindari kerumunan, menjaga kebersihan tangan, dan memakai masker. Kampanye kesehatan di tempat kerja atau lingkungan dapat menjadi solusi. Sedangkan pada perempuan di *cluster_2*, sosialisasi di posyandu atau pengajian bisa membantu ibu-ibu menjaga kesehatan pernapasan keluarga dan mencegah penyebaran penyakit.
- Puskesmas dapat bekerja sama dengan pemerintah Kelurahan Cigugur Tengah untuk memperbanyak program pencegahan, seperti penyemprotan disinfektan dan pembagian masker. Di tingkat RT dan RW, fokuskan upaya pencegahan di RT 10 (*cluster_0*), RT 1 (*cluster_1* dan *cluster_2*), serta RW 13, 14, dan 3 dengan membentuk kader kesehatan untuk sosialisasi dan pemantauan kesehatan warga.
- Program skrining di puskesmas penting untuk mendeteksi gejala pernapasan pada pasien baru. Pemeriksaan suhu tubuh dan edukasi tentang jarak dan etika batuk dapat mencegah penularan lebih lanjut, serta memastikan pasien mendapatkan penanganan yang tepat.
- Pelibatan tokoh masyarakat dan saluran komunikasi lokal seperti masjid atau balai warga sangat penting untuk mengedukasi masyarakat. Pembagian masker, vitamin, dan bahan edukasi dapat meningkatkan partisipasi masyarakat dalam mengendalikan penyebaran penyakit.

4. Kesimpulan

Penelitian ini menyimpulkan bahwa pengelompokan data pasien di Puskesmas Cigugur Tengah menggunakan algoritma *K-Means* dengan metode *elbow* menghasilkan cluster optimal sebanyak tiga ($k=3$). *Cluster_0* mencakup 268 data pasien dominan pria berusia 17–25 tahun dengan penyakit *Acute Nasopharyngitis* (J00) di wilayah RT 10, RW 13 dan 14. *Cluster_1* berisi 267 pasien dengan profil serupa namun dominan penyakit *Acute Upper Respiratory* (J06.9) di RT 1, RW 3. *Cluster_2* mencakup 268 data anak perempuan usia 0–5 tahun dengan penyakit dominan *Acute Nasopharyngitis* (J00) di RT 1, RW 14. Hasil performa menunjukkan rata-rata jarak ke *centroid* 6038,938 dan *Davies-Bouldin Index* sebesar 0,507, yang mengindikasikan kualitas pengelompokan yang optimal.

Penelitian ini dapat dikembangkan lebih lanjut dengan mempertimbangkan algoritma *clustering* lain serta parameter berbeda untuk meningkatkan akurasi hasil. Data yang lebih lengkap, seperti anamnesa pasien atau data dari dinas kesehatan tingkat kabupaten/provinsi, dapat meningkatkan spesifikasi dan cakupan penelitian. Selain itu, implementasi penelitian ke dalam sistem berbasis

desktop atau *web* disarankan untuk meningkatkan efisiensi dan kemudahan akses hasil analisis.

Ucapan Terimakasih

Saya mengucapkan terima kasih kepada Ketua Program Studi Teknik Informatika Politeknik TEDC Bandung atas bimbingan dan arahan selama proses pembelajaran dan penyusunan tugas akhir ini, Dosen Pembimbing atas kesabaran, waktu, dan tenaga yang diberikan selama proses bimbingan, Kepala TU Puskesmas Cigugur Tengah beserta staf atas izin dan dukungan dalam pelaksanaan penelitian, serta kepada kedua Orang Tua, Keluarga, dan Teman Dekat yang senantiasa memberikan bantuan, dukungan, dan semangat di setiap langkah perjalanan ini.

Daftar Pustaka:

- [1] Dinas Kesehatan Kota Cimahi, Profil Dinkes Cimahi, Cimahi: Dinas Kesehatan Kota Cimahi, 2022.
- [2] Rahayu, P., Sudipa, I. G. I., Suryani, Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M., Slamet, I., Harlina, S., & May Sanjaya, I. M, Buku Ajar Data Mining Vol. 1, Indonesia: PT. Sonpedia Publishing, 2024.
- [3] Gustientiedina, G., Adiya, M. H., & Desnelita, Y. "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan". *Jurnal Nasional Teknologi Dan Sistem Informasi*, vol 5 no 1, pp. 17–24, 2019.
- [4] Rosdianti, E., & Sugianto, C., "Analisis Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan (SPP) dengan Menggunakan Algoritma K-Means (Studi Kasus Politeknik TEDC Bandung)". *Journal Of Informatics And Electronics Engineering*, vol 3, no 1, pp. 25-32, 2023. doi:10.70428/jiee.v3i1.1063
- [5] Sugianto, C. A., Rahayu, A. H., Gusman, A., "Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah", *JOINT (Jurnal of Information Technology)*, vol.2, no.2, pp. 39-44, 2020
- [6] Olivia. L., "Pengelompokan Data Rekam Medis Untuk Mengetahui Penyakit Endemi Di Suatu Daerah Menggunakan *K-Means Clustering*". Repositori Institusi Universitas Sumatera Utara, 2019, [Online]. Tersedia: <https://repositori.usu.ac.id/handle/123456789/20892> [25 Desember 2024]
- [7] Jollyta, D., Efendi, S., Zarlis, M., & Mawengkang, H., "Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index", *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 918. 2019
- [8] Safira, R., & Agus, C., "Optimalisasi Algoritma K-Means Menggunakan Metode Elbow Dalam Pengelompokan Data Stunting", *Jurnal Brahmana*, vol. 5, no. 2, pp. 257–264. 2024.
- [9] Muningsih, E. "Optimalisasi Jumlah Cluster K-Means dengan Metode Elbow Untuk Pemetaan Pelanggan. Yogyakarta", *Prosiding Seminar Nasional Elinvo*, pp105 - 114, 14 September 2017.
- [10] Rahmayani, "Analisis clustering tingkat keparahan penyakit pasien menggunakan algoritma k-means, *Jurnal Inovasi Teknik Informatika*", vol. 1, no.2, pp 1-5, september 2018