

Implementasi Algoritma *K-Means* untuk *Clustering* Penyebaran Tuberkulosis di Kabupaten Karawang

Yanti Puspita Sari¹, Aji Primajaya², Agung Susilo Yuda Irawan³
Universitas Singaperbangsa Karawang¹²³
Jl. HS. Ronggowaluyo, Telukjambe Timur, Karawang, Jawa Barat, Indonesia
Email: yanti.16226@student.unsika.ac.id¹, aji.primajaya@staff.unsika.ac.id²,
agung@unsika.ac.id³

Abstrack – Tuberculosis is a health problem that is transmitted through the air due to the presence of the mycobacterium tuberculosis. Tuberculosis cases were spread almost throughout Karawang Regency. The area of Karawang Regency and the number of cases that tends to increase every year, allows the need for clustering areas of the tuberculosis spreads. The method that can be used to clustering the areas where tuberculosis is spreading is data mining. The data mining methodology applied is CRISP-DM. The dataset used comes from the Karawang District Health Office in 2018. Then the data is processed using the k-means clustering algorithm with the help of WEKA tools. The study begins by finding the best number of clusters using the elbow method, where K=3 is the best number of clusters. K=3 combined with S (random seed) 5, 10, 15 to be tested. The results of these testing then evaluated using SSE (Sum of Square Error). The best clustering is produced by K=3 with a combination of S=10 which consists of cluster 0 (7 members), cluster 1 (9 members), and cluster 2 (14 members). Evaluation values generated by SSE 2,4402 and Silhouette 0,5629.

Keywords – Tuberculosis, Data Mining, Clustering, CRISP-DM, K-Means

Intisari – Tuberkulosis merupakan masalah kesehatan yang sifatnya menular melewati udara karena adanya kuman *mycobacterium tuberculosis*. Kasus penyakit tuberkulosis tersebar hampir diseluruh daerah Kabupaten Karawang. Luasnya daerah Kabupaten Karawang dan jumlah kasus yang cenderung terus meningkat setiap tahunnya, memungkinkan perlunya pengelompokan daerah penyebaran penyakit tuberkulosis. Metode yang dapat digunakan untuk mengelompokkan daerah penyebaran penyakit tuberkulosis adalah *data mining*. Metodologi *data mining* yang diterapkan yaitu CRISP-DM. *Dataset* yang digunakan bersumber dari Dinas Kesehatan Kabupaten Karawang tahun 2018. Kemudian data tersebut diproses menggunakan algoritma *k-means clustering* dengan bantuan *tools* WEKA. Penelitian diawali dengan mencari jumlah *cluster* terbaik menggunakan metode *elbow*, dimana K=3 adalah jumlah *cluster* terbaik. K=3 dikombinasikan dengan S (*random seed*) 5, 10, 15 untuk diuji coba. Hasil dari uji coba tersebut kemudian dievaluasi menggunakan SSE (*Sum of Square Error*) dan *Silhouette*. *Clustering* terbaik dihasilkan oleh K=3 dengan kombinasi S=10 yang terdiri dari *cluster* 0 (7 anggota), *cluster* 1 (9 anggota), dan *cluster* 2 (14 anggota). Nilai evaluasi yang dihasilkan SSE 2,4402 dan *Silhouette* 0,5629.

Kata Kunci – Tuberkulosis, *Data Mining*, *Clustering*, CRISP-DM, *K-Means*

I. PENDAHULUAN

Tuberculosis ialah masalah kesehatan yang bersifat menular lewat udara, dimana sumber utama penularannya bisa dari batuk penderita tuberkulosis yang mengandung kuman berbahaya yaitu *mycobacterium tuberculosis* [1]. Dari 10 daftar penyebab kematian tertinggi di dunia, penyakit tuberkulosis masih menjadi salah satu penyebabnya. Angka kematian akibat penyakit ini secara keseluruhan diduga mencapai 1,3 juta jiwa [2]. Negara Indonesia menduduki *ranking* ke-2 secara global yang memiliki beban tuberkulosis tertinggi [3].

Kasus penyakit tuberkulosis di Indonesia hampir tersebar di semua daerah, salah satunya yaitu di Kabupaten Karawang. Pada tahun 2015 ditemukan kasus tuberkulosis di Kabupaten Karawang sebanyak 2.617 kasus. Dua tahun kemudian jumlahnya mengalami penurunan

menjadi 1.821 kasus, dan sangat disayangkan pada tahun 2018 jumlahnya kembali meningkat menjadi 2.075 kasus [4].

Dengan adanya permasalahan tersebut, memungkinkan perlunya suatu upaya penanganan yang efektif. Upaya tersebut salah satunya yaitu, dengan mengelompokkan daerah penyebaran penyakit tuberkulosis di Kabupaten Karawang. Pengelompokan ini bertujuan untuk mengetahui daerah mana saja yang memiliki kasus tuberkulosis tingkat tinggi, sedang, hingga rendah. Hasil dari penelitian ini diharapkan dapat menjadi bahan rekomendasi untuk Pemerintah Kabupaten Karawang dalam melakukan penanganan yang lebih efektif terhadap penyakit tuberkulosis. Pemerintah nantinya dapat lebih fokus terlebih dahulu ke daerah-daerah yang memiliki kasus tuberkulosis tingkat tinggi, sehingga jumlah kasus tuberkulosis di Kabupaten Karawang dapat semakin berkurang.

Untuk mengelompokkan daerah penyebaran penyakit tuberkulosis dapat menggunakan metode *clustering*. Cara kerja metode *clustering* yaitu dengan menggabungkan objek data yang memiliki kesamaan sifat atau karakteristik kedalam satu kelompok (*cluster*) kecil [5]. Dalam implementasinya metode *clustering* harus melakukan tahapan algoritma agar tujuannya mampu tercapai. Algoritma *k-means* dapat diimplementasikan pada metode *clustering* [6]. Tugas dari *k-means* adalah membagi data menjadi beberapa kelompok berdasarkan tingkat kemiripannya [7].

II. SIGNIFIKANSI STUDI

A. Studi Literatur

Pada penelitian ini, penulis mengumpulkan beberapa literatur yang berhubungan dengan penelitian yang dilakukan. Adapun studi literaturnya yaitu sebagai berikut:

1. Tuberkulosis

Tuberkulosis merupakan masalah kesehatan yang bersifat menular lewat udara, dimana sumber utama penularannya bisa dari batuk penderita tuberkulosis yang mengandung kuman berbahaya yaitu *Mycobacterium Tuberculosis* [1]. Berat badan menurun, nyeri dada, meriang, sesak napas, batuk yang disertai dahak dan darah, merupakan tanda-tanda seseorang menderita tuberkulosis [3].

2. Data Mining

Kegiatan mencari *knowledge* terpendam dari himpunan data yang amat banyak adalah definisi dari *data mining* [8]. *Output* atau hasil dari proses *data mining* bisa dijadikan bahan untuk menentukan langkah di waktu yang akan datang [9]. Berdasarkan tugasnya *data mining* terbagi menjadi 6 metode yaitu *description*, *estimation*, *prediction*, *classification*, *clustering*, serta *association* [10]. *Data mining* berfungsi untuk menambang pengetahuan berharga yang tersembunyi di dalam sebuah data [11].

3. Clustering

Clustering bekerja dengan cara menggabungkan objek data yang memiliki kesamaan sifat atau karakteristik tertentu kedalam satu kelompok (*cluster*) kecil [5]. Dalam *clustering* dipastikan masing-masing anggota dari satu *cluster* mempunyai kemiripan satu sama lain, dan sebaliknya antara anggota *cluster* satu dengan anggota *cluster* lain tidak akan sama [12].

4. K-Means

K-Means didefinisikan sebagai algoritma yang bertugas membagi data menjadi kelompok berdasarkan tingkat kemiripannya [7]. Keunggulan *k-means* dibandingkan algoritma yang lainnya adalah lebih sederhana, mudah diimplementasikan, tidak lambat, mudah disesuaikan, serta paling sering digunakan dalam proses *data mining* [13].

Tahapan algoritma *k-means* yaitu [14] :

- 1) Memutuskan berapa banyak *cluster* yang ingin dibentuk.
- 2) Memilih *centroid* dengan *random* sebanyak angka *cluster*.
- 3) Hitunglah *distance* terhadap *centroid* menggunakan *euclidean distance*. Adapun rumusnya sebagai berikut:

$$J(a_i, b_i) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \tag{1}$$

Keterangan:

$J(a_i, b_i)$: jarak objek antara nilai data dan *centroid*

a_i : nilai data dari dimensi

b_i : nilai *centroid* dari dimensi

n : banyaknya dimensi atau atribut data

4) Gabungkan data berdasarkan jarak terpendek dengan *centroid*.

5) Tentukan nilai *centroid* baru menggunakan rumus berikut:

$$W_k = \frac{1}{n_k} \sum j_i \tag{2}$$

Keterangan:

n_k : jumlah data dalam *cluster*

j_i : jumlah nilai jarak yang masuk dalam setiap *cluster*

6) Ulangilah tahap ke-3 sampai 5 sampai keanggotaan setiap *cluster* tetap.

5. Silhouette

Silhouette adalah salah satu indeks validitas untuk mengevaluasi hasil algoritma *clustering*. *Silhouette* menghitung rata-rata nilai dari setiap titik data. Perhitungan ini merupakan selisih nilai *separation* dan *compactness* yang dibagi dengan maksimum antara keduanya. Hasil *cluster* yang baik ditunjukkan dengan nilai *silhouette* yang mendekati rentang nilai 1 [6]. Berikut ini merupakan rentang nilai *silhouette* dan interpretasinya:

TABEL I
RENTANG NILAI *SILHOUETTE* DAN INTERPRETASI

Rentang Nilai	Interpretasi
0.71 – 1.0	Struktur yang kuat
0.51 – 0.70	Struktur yang beralasan
0.26 – 0.50	Struktur lemah
< 0.25	Tidak ditemukan struktur yang substansial

6. CRISP-DM

CRISP-DM atau *Cross-Industry Standard Process for Data Mining* adalah suatu metode terdiri dari beberapa tahapan analisis yang digunakan untuk menyelesaikan permasalahan *data mining*. CRISP-DM memiliki 6 tahapan analisis, yaitu *business understanding* (pemahaman bisnis), *data understanding* (pemahaman data), *data preparation* (persiapan data), *modeling* (pemodelan), *evaluation* (evaluasi), dan *deployment* (penyebaran) [15].

B. Data

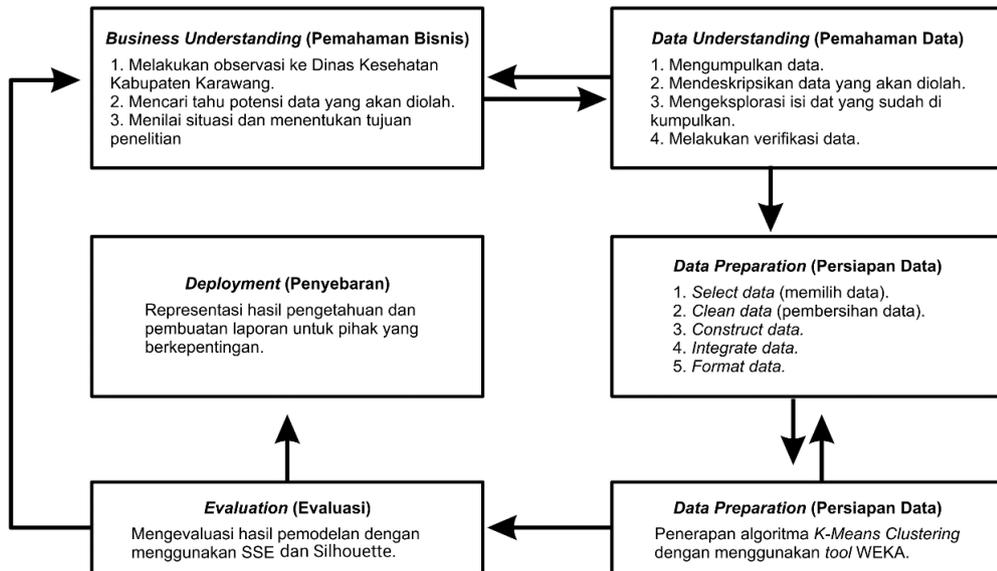
Data yang digunakan pada penelitian ini adalah data kasus tuberkulosis di Karawang tahun 2018. Data tersebut berjumlah 30 data dan terdiri dari beberapa variabel seperti no, kecamatan, puskesmas, jumlah penduduk, kasus baru TBC, seluruh kasus TBC, dan kasus TBC anak.

C. Lokasi Penelitian

Dinas Kesehatan Kabupaten Karawang merupakan lokasi pada penelitian ini. Dinas tersebut beralamat di Jalan Parahyangan No. 39, Adiarsa Barat, Kecamatan Karawang Barat, Kabupaten Karawang, Jawa Barat.

D. Metode Penelitian dan Evaluasi

Penelitian ini hanya mengimplementasi algoritma *k-means* untuk *clustering* data kasus tuberkulosis di Kabupaten Karawang pada tahun 2018 tanpa pembuatan sebuah program. Data tersebut bersumber dari Dinas Kesehatan Kabupaten Karawang. Penentuan jumlah *cluster* yang akan dibentuk menggunakan metode *elbow*, hal ini bertujuan agar jumlah *cluster* lebih optimal. *Dataset* yang telah diperoleh akan diolah menggunakan *tools* WEKA (*Waikato Environment for Knowledge Analysis*). Adapun tahapan dari penelitian yang dilakukan dengan menggunakan metodologi CRISP-DM adalah sebagai berikut:



Gambar 1. Rancangan Penelitian

Gambar 1 adalah rancangan penelitian yang dilakukan dengan metodologi CRISP-DM. Tahap *business understanding* merupakan proses observasi ketempat penelitian yaitu Dinas Kesehatan Kabupaten Karawang, guna mencari tahu potensi data yang akan diolah. *Data understanding* terdiri dari kegiatan mengumpulkan data, mendeskripsikan dan mengeksplorasi isi data, serta melakukan verifikasi kualitas data. Tahap *data preparation* yang dilakukan yaitu memilih data yang akan digunakan (data kasus tuberkulosis di Kabupaten Karawang tahun 2018), membersihkan data dari *missing value*, dan juga mengubah data kedalam *range* 0-1 menggunakan *mix-max normalization*. Pada tahap *modeling* algoritma *k-means* akan diterapkan untuk mengelompokkan daerah penyebaran penyakit tuberkulosis di Kabupaten Karawang. Hasil *clustering* akan dievaluasi menggunakan SSE (*Sum of Square Error*) serta *Silhouette*. Tahap terakhir yaitu *deployment* merupakan bentuk laporan mengenai pengetahuan yang didapat dari data yang telah selesai diolah.

III. HASIL DAN PEMBAHASAN

A. Business Understanding (Pemahaman Bisnis)

Tahap *business understanding* adalah proses menentukan tujuan dilakukannya penelitian. Tujuan dari penelitian ini yaitu mengelompokkan daerah penyebaran tuberkulosis di Karawang menggunakan algoritma *k-means*. Kegiatan observasi dilakukan secara langsung ke Dinas Kesehatan Kabupaten Karawang guna mencari tahu potensi data yang akan diolah. Hasil dari penelitian ini diharapkan bisa bermanfaat untuk Pemerintah Karawang dalam melakukan upaya penanganan penyakit tuberkulosis.

B. Data Understanding (Pemahaman Data)

Setelah menentukan tujuan penelitian tahap selanjutnya adalah *data understanding* atau pemahaman data. *Dataset* awal yang digunakan pada penelitian ini yaitu data kasus tuberkulosis di Kabupaten Karawang pada tahun 2018. Kemudian data awal akan dideskripsikan dan dieksplorasi, serta diverifikasi kualitasnya. Berikut ini merupakan tabel hasil *describe the data* dan *explore the data*:

TABEL II
DESKRIPSI DATA KASUS TUBERKULOSIS

Atribut	Tipe Data	Keterangan
No	Numerik	Nomor urut daftar Kecamatan
Kecamatan	Nominal	Kecamatan di Karawang
Puskesmas	Nominal	Puskesmas di Karawang
Jumlah Penduduk	Numerik	Jumlah Penduduk Kabupaten Karawang
Kasus Baru TBC	Numerik	Total kasus baru penyakit <i>tuberculosis</i> BTA+ di suatu Kecamatan dan Puskesmas
Seluruh Kasus TBC	Numerik	Total seluruh kasus <i>tuberculosis</i> di suatu Kecamatan dan Puskesmas
Kasus TBC Anak	Numerik	Total kasus <i>tuberculosis</i> anak usia 0-14 tahun di suatu Kecamatan dan Puskesmas

Tabel II merupakan deskripsi data kasus tuberkulosis di Kabupaten Karawang pada tahun 2018. *Describe the data* bertugas mendeskripsikan atribut, tipe data, dan keterangan atribut data yang telah didapat dari Dinas Kesehatan Kabupaten Karawang.

TABEL III
EXPLORE DATA KASUS TUBERKULOSIS

Atribut	Tipe Data	Keterangan
No	Numerik	Nomor urut daftar Kecamatan Tahun 2018: 1-30
Kecamatan	Nominal	Tahun 2018: Terdapat 30 Kecamatan Nama Kecamatan yang ada di Karawang
Puskesmas	Nominal	Tahun 2018: Terdapat 30 Kecamatan Nama Puskesmas yang ada di Karawang
Jumlah Penduduk	Numerik	Tahun 2018: Terdapat 50 Puskesmas Jumlah Penduduk Kabupaten Karawang
Kasus Baru TBC	Numerik	Tahun 2018: 2.336.009 jiwa Total kasus baru penyakit <i>tuberculosis</i> BTA+ di suatu Kecamatan dan Puskesmas
Seluruh Kasus TBC	Numerik	Tahun 2018: 1.306 kasus Total seluruh kasus <i>tuberculosis</i> di suatu Kecamatan dan Puskesmas
Kasus TBC Anak	Numerik	Tahun 2018: 2.075 kasus Total kasus <i>tuberculosis</i> anak usia 0-14 tahun di suatu Kecamatan dan Puskesmas
		Tahun 2018: 138 kasus

Tabel III merupakan *explore the data* yang bertugas untuk memperoleh pengetahuan dari karakteristik data yang digunakan, sehingga data lebih mudah untuk dipahami.

Tahap terakhir dari proses *data understanding* (pemahaman data) adalah memverifikasi kualitas data. Kualitas data dikatakan baik apabila tidak terdapat nilai yang hilang dan duplikasi. Pada data kasus tuberkulosis tidak ditemukan *missing value* (semua kolom memiliki nilai) dan juga tidak ditemukan duplikasi data.

C. Data Preparation (Persiapan Data)

Data preparation merupakan semua aktivitas pengolahan data awal yang bertujuan untuk membuat *dataset final*. Dalam penelitian ini, data kasus tuberkulosis akan dibersihkan terlebih dahulu dari atribut data yang tidak diperlukan saat proses pemodelan. Selanjutnya data akan dinormalisasikan kedalam *range 0-1* menggunakan *min-max normalization*. Dalam proses perhitungan normalisasi data, *tools* yang digunakan yaitu *Microsoft Excel*. Berikut adalah tabel dari data kasus tuberkulosis setelah dilakukan pembersihan atribut dan normalisasi data:

TABEL IV
DATA KASUS TUBERKULOSIS SETELAH *CLEANING*

No	Kecamatan	Jumlah Penduduk	Kasus Baru TBC	Seluruh Kasus TBC	Kasus TBC Anak
1.	Pangkalan	41.723	17	35	0
2.	Tegalwaru	39.847	30	51	2
3.	Ciampel	39.444	7	51	5
4.	Telukjambe Timur	116.110	74	99	0
5.	Telukjambe Barat	54.323	98	98	22
6.	Klari	145.495	139	154	5
7.	Cikampek	112.354	223	230	35
8.	Purwasari	61.821	72	86	0
9.	Kotabaru	117.454	82	155	14
10.	Tirtamulya	52.885	5	9	0
11.	Jatisari	80.284	30	55	4
...
30.	Pakisjaya	44.876	2	6	0

Tabel IV merupakan data yang sudah melalui proses *cleaning*, dimana atribut data kasus tuberkulosis di Kabupaten Karawang pada tahun 2018 setelah proses *cleaning* tersisa 5 atribut. Sedangkan satu atribut Puskesmas telah dihapus, karena atribut tersebut tidak diperlukan saat proses pemodelan menggunakan algoritma *k-means*.

TABEL V
DATA KASUS TUBERKULOSIS SETELAH NORMALISASI

No	Kecamatan	Jumlah Penduduk	Kasus Baru TBC	Seluruh Kasus TBC	Kasus TBC Anak
1.	Pangkalan	0,020414379	0,076233184	0,152173913	0
2.	Tegalwaru	0,003609914	0,134529148	0,22173913	0,057142857
3.	Ciampel	0	0,031390135	0,22173913	0,142857143
4.	Telukjambe Timur	0,686743642	0,331838565	0,430434783	0
5.	Telukjambe Barat	0,133280185	0,439461883	0,426086957	0,628571429
6.	Klari	0,949962826	0,623318386	0,669565217	0,142857143
7.	Cikampek	0,653098883	1	1	1
8.	Purwasari	0,200444297	0,322869955	0,373913043	0
9.	Kotabaru	0,698782662	0,367713004	0,673913043	0,4
10.	Tirtamulya	0,120399151	0,022421525	0,039130435	0
11.	Jatisari	0,365828534	0,134529148	0,239130435	0,114285714

...
30.	Pakisjaya	0,048657703	0,00896861	0,026086957	0

Tabel V merupakan data yang telah melalui proses normalisasi *min-max*. Normalisasi data dilakukan karena adanya perbedaan *range* nilai yang cukup jauh antara atribut jumlah penduduk dengan atribut lain. Perbedaan *range* tersebut dikhawatirkan dapat menyebabkan hasil akurasi dari proses pemodelan menjadi kurang baik.

D. Modeling (Pemodelan)

Pada tahap *modeling* (pemodelan) algoritma *k-means clustering* akan diimplementasikan. Dari hasil pemrosesan data kasus tuberkulosis tahun 2018 menggunakan metode *elbow* telah didapat K (jumlah *cluster*) terbaik, yaitu K=3. Pengklasterannya K=3 akan dikombinasikan dengan S (*random seed*) S=5, S=10, dan S=15 menggunakan *tools* WEKA 3.8.3. Berikut ini adalah hasil *clustering* kombinasi K=3 dengan *random seed* 5, 10, dan 15:

```

kMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 2.5342732792863036

Initial starting points (random):

Cluster 0: 0.63625,0.206278,0.513043,0.114286
Cluster 1: 0.199217,0.098655,0.208696,0.028571
Cluster 2: 0.050602,0.013453,0.021739,0
    
```

Gambar 2. Hasil *Clustering* K=3 dengan S=5

Pada gambar 2 menunjukkan hasil *clustering* data kasus penyakit tuberkulosis di Karawang pada tahun 2018 dengan kombinasi K=3 dan S=5 melakukan iterasi sebanyak 3 kali dan menghasilkan SSE sebesar 2,5342732792863036.

```

kMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 2.4402041623523654

Initial starting points (random):

Cluster 0: 0.686744,0.331839,0.430435,0
Cluster 1: 0.200444,0.32287,0.373913,0
Cluster 2: 0.120399,0.022422,0.03913,0
    
```

Gambar 3. Hasil *Clustering* K=3 dengan S=10

Pada gambar 3 menunjukkan hasil *clustering* data kasus penyakit tuberkulosis di Karawang pada tahun 2018 dengan kombinasi K=3 dan S=10 melakukan iterasi sebanyak 3 kali dan menghasilkan nilai SSE (*Sum of Square Error*) sebesar 2,4402041623523654.

```

kMeans
=====

Number of iterations: 2
Within cluster sum of squared errors: 2.5163278543388388

Initial starting points (random):

Cluster 0: 0.242223,0.197309,0.230435,0
Cluster 1: 0.698783,0.367713,0.673913,0.4
Cluster 2: 0,0.03139,0.221739,0.142857
    
```

Gambar 4. Hasil *Clustering* K=3 dengan S=15

Pada gambar 4 menunjukkan hasil *clustering* data kasus penyakit tuberkulosis di Karawang pada tahun 2018 dengan kombinasi K=3 dan S=15 melakukan iterasi sebanyak 2 kali dan menghasilkan nilai SSE (*Sum of Square Error*) sebesar 2,5163278543388388.

E. *Evaluation (Evaluasi)*

Hasil *clustering* yang didapat dari jumlah *cluster* K=3 dengan kombinasi *random seed* 5, 10, dan 15 dievaluasi menggunakan SSE (*Sum of Square Error*). Hasil pengklusteran terbaik apabila mempunyai nilai SSE terkecil.

TABEL VI
HASIL SSE (*SUM OF SQUARE ERROR*)

Tahun	K	S=5	S=10	S=15
2018	3	2,5342732792863036	2,4402041623523654	2,5163278543388388

Berdasarkan tabel di atas menunjukkan nilai SSE yang paling rendah (hasil *clustering* terbaik) pada data kasus tuberkulosis tahun 2018 adalah hasil *clustering* dari kombinasi K=3 dengan *random seed* S=10 yang menghasilkan nilai SSE (*Sum of Square Error*) yaitu sebesar 2,4402041623523654 dan melakukan iterasi sebanyak 3 kali.

TABEL VII
HASIL EVALUASI *SILHOUETTE*

Tahun	K	<i>Silhouette</i>	Interpretasi
2018	3	0,5629	Struktur yang beralasan

Pada tabel VII menunjukkan hasil *clustering* dari data kasus penyakit tuberkulosis tahun 2018 dengan menggunakan algoritma *k-means* dan jumlah *cluster* 3 yang menghasilkan nilai evaluasi *Silhouette* sebesar 0,5629 (struktur yang beralasan).

F. *Deployment (Penyebaran)*

Deployment (penyebaran) merupakan representasi mengenai *knowledge* atau pengetahuan yang didapat dari data yang telah diolah dalam proses *data mining*. Rencana penyebaran pada penelitian ini yaitu penyerahan hasil penelitian kepada Dinas Kesehatan Kabupaten Karawang sebagai bahan evaluasi dalam melakukan upaya penanganan penyakit tuberkulosis, sehingga diharapkan jumlah kasus tuberkulosis di Kabupaten Karawang dapat berkurang. Berikut adalah hasil keanggotaan *cluster* pada data kasus tuberkulosis di Kabupaten Karawang tahun 2018:

TABEL VIII
 ANGGOTA *CLUSTER* DATA KASUS TUBERKULOSIS TAHUN 2018

Nama Kecamatan		
C0	C1	C2
Telukjambe Timur	Telukjambe Barat	Pangkalan
Klari	Purwasari	Tegalwaru
Cikampek	Jatisari	Ciampel
Kotabaru	Cilamaya Kulon	Tirtamulya
Karawang Barat	Rawamerta	Banyusari
Karawang Timur	Kutawaluya	Cilamaya Wetan
Rengasdengklok	Pedes	Lemahabang
	Cibuaya	Telagasari
	Batujaya	Majalaya
		Tempuran
		Jayakarta
		Tirtajaya
		Cilebar
		Pakisjaya

Pada tabel di atas merupakan daftar keanggotaan *cluster* 0, *cluster* 1, dan *cluster* 2 yang dihasilkan dari kombinasi terbaik yaitu $K=3$ dengan $S=10$ pada data kasus tuberkulosis tahun 2018 dengan nilai SSE sebesar 2,4402041623523654. Anggota pada *cluster* 0 terdiri dari 7 kecamatan, *cluster* 1 sebanyak 9 kecamatan, dan untuk *cluster* 2 sebanyak 14 kecamatan.

TABEL IX
 RATA-RATA KASUS TUBERKULOSIS TAHUN 2018

Cluster	Rata-rata Kasus Tuberkulosis Tahun 2018	Kategori
0	269,1428571	Tinggi
1	121,8889	Sedang
2	33,71428571	Rendah

Pada tabel IX menunjukkan data kasus tuberkulosis tahun 2018 terbagi menjadi tiga kategori yaitu daerah dengan rata-rata kasus tuberkulosis tinggi, sedang, hingga rendah. *Cluster* 0 merupakan daerah kategori tinggi dengan rata-rata kasusnya mencapai angka 269,1428571 kasus. Kemudian disusul oleh *cluster* 1 dengan kategori sedang, dimana rata-rata kasus penyakit tuberkulosis yang ada pada *cluster* tersebut yaitu sebesar 121,8889 kasus. Sedangkan untuk kategori daerah rendah yaitu ada pada *cluster* 2 dengan rata-rata 33,71428571 kasus.

IV. KESIMPULAN

Penelitian ini menerapkan algoritma *k-means* untuk *clustering* daerah penyebaran penyakit tuberkulosis di Karawang pada tahun 2018. Hasil *clustering* terbaik merupakan kombinasi dari jumlah *cluster* $K=3$ dengan *random seed* $S=10$ yang menghasilkan nilai evaluasi SSE sebesar 2,4402. Sedangkan hasil evaluasi dari *Silhouette* yaitu 0,5629 (struktur yang beralasan). *Cluster* 0 terdiri dari (7 kecamatan), *cluster* 1 (9 kecamatan), dan *cluster* 2 (14 kecamatan). Kemudian daerah dengan kategori kasus tuberkulosis tingkat tinggi yaitu daerah yang ada pada *cluster* 0 dengan keanggotaan *cluster* terdiri dari kecamatan Telukjambe Timur, Cikampek, Kotabaru, Klari, Karawang Barat, Karawang Timur, dan Rengasdengklok. Kecamatan tersebut nantinya

harus menjadi fokus pertama Pemerintah Karawang dalam upaya penanganan tuberkulosis. Adapun saran untuk penelitian selanjutnya yaitu diharapkan menggunakan algoritma *clustering* dan metode evaluasi yang berbeda, sehingga nantinya dapat dijadikan bahan perbandingan hasil algoritma dan metode evaluasi mana yang jauh lebih baik.

REFERENSI

- [1] Amrin and H. Saiyar, “Aplikasi Diagnosa Penyakit Tuberculosis Menggunakan Algoritma Data Mining,” *PARADIGMA*, vol. 20, no. 2, pp. 91–97, 2018.
- [2] Kementerian Kesehatan RI, *Profil Kesehatan Indonesia Tahun 2017*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2018.
- [3] D. Febriyanto and Y. I. Kurniawan, “Prediksi Penyakit Tuberculosis (TBC) Menggunakan Algoritma C4.5,” *J. Ilm. Sinus*, vol. 16, no. 2, pp. 23–36, 2018.
- [4] Dinas Kesehatan Kabupaten Karawang, *Profil Kesehatan Kabupaten Karawang Tahun 2018*. Karawang: Dinas Kesehatan Kabupaten Karawang, 2019.
- [5] Asroni and R. Adrian, “Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang,” *J. Ilm. Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2015.
- [6] A. Aditya, I. Jovian, and B. N. Sari, “Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018 / 2019,” *J. Media Inform. Budidarma*, vol. 4, no. 1, pp. 51–58, 2020.
- [7] N. Butarbutar, A. P. Windarto, D. Hartama, and Solikhun, “Komparasi Kinerja Algoritma Fuzzy C-Means dan K-Means dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai Akademik Siswa (Studi Kasus: SMP Negeri 2 Pematangsiantar),” *JURASIK (Jurnal Ris. Sist. Inf. Tek. Inform.)*, vol. 1, no. 1, pp. 46–55, 2016.
- [8] R. Setiawan, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik LP3I Jakarta),” *J. Lentera ICT*, vol. 3, no. 1, pp. 76–92, 2016.
- [9] M. G. Sadewo, A. P. Windarto, and A. Wanto, “Penerapan Algoritma Clustering dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/Mitigasi Bencana Alam Menurut Provinsi dengan K-Means,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018.
- [10] Y. Mardi, “Data Mining: Klasifikasi Menggunakan Algoritma C4.5,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [11] A. F. Lestari and M. Hafiz, “Penerapan Algoritma Apriori Pada Data Penjualan Barbar Warehouse,” *J. INOVTEK Polbeng Seri Inform.*, vol. 5, no. 1, pp. 96–105, 2020.
- [12] H. Sulastris and A. I. Gufroni, “Penerapan Data Mining dalam Pengelompokan Penderita Thalassaemia,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017.
- [13] A. Ramadhan, Z. Efendi, and Mustakim, “Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling,” *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, pp. 219–226, 2017.
- [14] A. T. Rahman, Wiranto, and R. Anggrainingsih, “Coal Trade Data Clustering Using K-Means (Case Study PT. Global Bangkit Utama),” *ITSMART J. Ilm. Teknol. dan Inf.*, vol. 6, no. 1, pp. 24–31, 2017.
- [15] B. E. Adiana, I. Soesanti, and A. E. Permanasari, “Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering,” *JUTEI*, vol. 2, no. 1, pp. 23–32, 2018.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Dinas Kesehatan Kabupaten Karawang yang telah memberikan izin penelitian, sehingga penulis dapat dengan mudah dalam memperoleh

data yang dibutuhkan. Demikianlah semoga banyak manfaat yang dapat diambil bagi semua pihak.