



Penerapan K-Means Clustering dalam Pengelompokan Kasus Tuberkulosis di Provinsi Jawa Barat

Fadhlan Sulistiyo Hidayat^{1*}, Rizma Berliana Putri Affandi², Virgaria Zuliana³,
Tesa Nur Padilah⁴

^{1,2,3}Universitas Singaperbangsa Karawang

Received: 10 Agustus 2022

Revised: 14 Agustus 2022

Accepted: 18 Agustus 2022

Abstract

Tuberculosis is a very common infectious disease and is lethal in most of the cases. This is the background of this research, namely because there are many cases of Tuberculosis in Jawa Barat. According to data obtained from the Open Data Jabar, namely Tuberculosis Data in Jawa Barat Province, showing data that in 2020 all districts and cities in Jawa Barat had a number of Tuberculosis cases starting from 320 cases in Banjar Regency which was the lowest case, and 10,248 cases in Bogor Regency which is the highest case in Jawa Barat. The purpose of this study was to cluster TB cases into high and low categories based on gender. The data we use is data on the number of TB cases in Jawa Barat province in 2020 which consists of 27 districts/cities. In this study using the Clustering method with the K-Means algorithm. The results obtained based on the test, the clusters obtained were 2 with cluster 0 with 23 low TB cases and 4 clusters for high TB cases. Researchers hope that the results of this study can become knowledge for the government to reduce the number of TB in Jawa Barat.

Keywords: Clustering; K-Means; Data Mining; Tuberculosis

(*) Corresponding Author: fadhlan.sulistiyo18061@student.unsika.ac.id

How to Cite: Hidayat, F., Affandi, R. B., Zuliana, V., & Padilah, T. (2022). Application of K-Means Clustering in Grouping Tuberculosis Cases in West Java Province. *Jurnal Ilmiah Wahana Pendidikan*, 8(15), 213-227. <https://doi.org/10.5281/zenodo.7049113>

INTRODUCTION

Tuberkulosis atau yang sering disingkat dengan TBC atau TB (*Tubercle Bacillus*) merupakan infeksi yang sangat umum dan berpotensi besar untuk membunuh atau mematikan dalam sebagian besar kasusnya. Pengidap tuberkulosis mengalami berbagai gejala, termasuk demam, kurus, dan batuk. Meskipun pengobatan yang berhasil mencegah kematian, banyak penderita tuberkulosis yang mengalami masalah kesehatan yang berkelanjutan setelah episode penyakit, dan ada peningkatan bukti kecacatan jangka panjang dan peningkatan risiko kematian pada populasi ini[1]. Tuberkulosis memiliki keterkaitan antara manusia dengan lingkungannya, terutama di daerah perkotaan yang memiliki populasi dan kepadatan tertinggi sehingga informasi yang akurat tentang lingkungan perkotaan daerah Tuberkulosis itu penting[2].

Berdasarkan hal tersebut, yang melatarbelakangi proses riset ini adalah karena banyaknya masalah Tuberkulosis di wilayah Jawa Barat. Menurut data yang diperoleh dari Open Data Jabar[3] yaitu Data Tuberkulosis di Provinsi Jawa Barat menampilkan data bahwa pada tahun 2020 seluruh kota dan kabupaten di wilayah Jawa Barat memiliki angka kasus Tuberkulosis sebanyak mulai dari angka 320 kasus di Kabupaten Banjar yang merupakan kasus terendah, dan 10,248 kasus di Kabupaten Bogor yang merupakan kasus tertinggi di Jawa Barat.

Banyaknya kasus Tuberkulosis dapat berbahaya bagi seluruh masyarakat di Indonesia khususnya di Jawa Barat, maka dari itu diperlukan pencegahan terhadap kasus



dari penyakit tersebut yang salah satunya dapat menggunakan teknik *data mining* untuk menganalisis *cluster* tinggi dan rendah pada suatu daerah.

Cara kerja dari teknik *clustering* yaitu populasi atau beberapa data yang kita punya dikelompokkan di dalam beberapa *cluster* sehingga titik data dalam kelompok yang sama lebih mirip dengan titik data lain dalam kelompok yang sama daripada yang ada di kelompok lain[4]. Pada penelitian ini peneliti bertujuan untuk mengetahui *cluster* tinggi dan rendah sehingga nantinya hasil riset ini dapat dijadikan materi evaluasi untuk Dinas Kesehatan Jawa Barat untuk dilakukan pencegahan terhadap penyakit Tuberkulosis khususnya di Jawa Barat.

Penelitian terdahulu berjudul Penerapan Algoritma *K-Means Clustering* untuk Mendeteksi Penyebaran Penyakit TBC (Studi Kasus: Di Kabupaten Deli Serdang) dengan algoritma *K-Means Clustering* membahas tentang pengelompokan daerah penyebaran penyakit diare di Kabupaten Deli Serdang dengan menggunakan *dataset* penyebaran diare di Kabupaten Langkat dengan jumlah 22 kecamatan dengan tujuan untuk mencari titik-titik pusat *centroid* daerah penyebaran penyakit Deli Serdang dengan hasil dapat membuat titik *cluster* penyebaran penyakit TBC[4].

Penelitian lainnya juga menerapkan algoritma yang sama dengan judul Penerapan Metode *K-Means Clustering* Data COVID-19 di Provinsi Jakarta yang membahas tentang pengelompokan *clustering* dalam menangani kasus COVID-19 dengan 2 *cluster* yaitu *cluster* dengan kasus positif dan tingkat kesembuhan tertinggi maupun terendah. Hasil dari penelitian tersebut didapatkan 2 *cluster* dengan atribut positif dan sembuh. *Cluster* tersebut menjelaskan tentang *cluster* tinggi dan rendah di daerah Jakarta[5].

Berdasarkan riset yang sudah dilakukan oleh peneliti sebelumnya, riset atau penelitian sekarang dilakukan dengan tujuan untuk mengetahui pengaruh jenis kelamin terhadap masalah tuberkulosis di wilayah Jawa Barat serta klasterisasi berdasarkan *cluster* tinggi dan *cluster* rendah di beberapa kabupaten dan kota di wilayah Jawa Barat.

Penelitian ini menggunakan data kasus Tuberkulosis yang ada di Jawa Barat yang merupakan provinsi dengan jumlah populasi terbanyak sehingga bisa menjadi tempat penyebaran kasus Tuberkulosis yang tinggi. Selain menggunakan data kasus yang ada di Jawa Barat, penelitian ini juga menggunakan atribut yang berbeda dengan penelitian lain yaitu atribut jenis kelamin untuk mengetahui pengaruhnya terhadap banyaknya kasus Tuberkulosis.

METHODS

a. Data Cleaning

Data Cleaning adalah proses menghapus atau memperbaiki data yang salah, tidak lengkap, salah format, rusak atau duplikat dalam sekumpulan data. Saat menyatukan beberapa sumber data, ada banyak peluang untuk data diduplikasi atau diberi label yang salah. *Data Cleaning* atau Pembersihan data adalah proses memastikan bahwa data yang dikumpulkan benar, konsisten, dan dapat digunakan. Data tersebut kemudian disiapkan untuk dianalisis dengan menghapus atau memodifikasi data yang tidak benar, tidak lengkap, tidak relevan, duplikat, atau tidak diformat dengan benar. Proses ini bertujuan untuk mendeteksi dan mengoreksi (atau menghilangkan) *records* yang rusak atau tidak akurat dari *record set*, tabel, atau *database* sehingga data tersebut bisa digunakan untuk proses permodelan *data mining*[6].

b. Data Selection & Transformation

Data Selection didefinisikan sebagai proses penentuan jenis dan sumber data yang sesuai, serta instrumen yang sesuai untuk mengumpulkan data. *Data Selection* atau

seleksi data mencari data yang relevan atau berhubungan dengan analisis yang akan diterima dari pengumpulan data. Pada tahap ini diputuskan untuk mengambil data mana yang akan digunakan sebagai input untuk metode *data mining*[7]. *Data Transformation* adalah proses mengubah format, struktur, atau nilai data sehingga data tersebut bisa dimodelkan dengan atribut lainnya.

c. Data Mining

Data mining adalah kegiatan penambangan data dengan proses yang ditujukan untuk analisis basis data yang sangat besar untuk mengekstrak informasi dan pengetahuan yang mungkin terbukti akurat dan berpotensi berguna bagi *Data Scientist* yang terlibat dalam pengambilan dan pemecahan masalah[8]. *Data mining* mengacu pada penggalian atau penambangan pengetahuan dari sejumlah besar data. Hal ini juga didefinisikan sebagai menemukan informasi tersembunyi dari database. Ini adalah teknik yang digunakan terutama untuk menemukan pola yang tidak diketahui dan yang mengubah data mentah menjadi informasi yang dapat dimengerti pengguna[9].

d. Clustering

Clustering adalah salah satu metode dalam *data mining* yang bertujuan untuk menganalisa suatu data atau informasi yang dapat menyelesaikan permasalahan pengelompokan data dari suatu dataset ke dalam subset data[10]. Metode ini adalah cara yang efisien untuk mengelompokkan data ke dalam kelas yang berbeda berdasarkan skema internal dan yang sebelumnya tidak diketahui yang melekat pada data[11].

e. K-Means

K-Means merupakan suatu metode *clustering* dengan membagi data ke dalam satu atau lebih cluster, data beserta karakteristik yang mirip akan dikelompokkan dalam *cluster* yang sama dan karakteristik lainnya akan dikelompokkan ke dalam kelompok yang sesuai dengan tingkat kemiripannya masing-masing[12]. *K-Means Clustering* merupakan metode yang berfungsi untuk melakukan analisa data atau metode yang mengerjakan proses pemodelan tanpa *supervise* dan salah satu cara untuk melakukan *clustering* data pada sistem partisi.

Tahapan Metode K-Means

Pada metode *K-Means* memiliki beberapa tahapan yang perlu dilakukan saat melakukan perhitungan untuk mencari data yang terklaster hingga iterasinya sama dengan perhitungan sebelumnya. Tahapan teknik *clustering* dengan metode *K-Means* adalah sebagai berikut.

- 1) Menentukan **k** sebagai jumlah *cluster* (C1, C2, dan C3) yang akan dibentuk.
- 2) Tentukan nilai untuk pusat *cluster*.
- 3) Gunakan rumus *Euclidean Distance* untuk menghitung jarak tiap data input terhadap semua pusat *cluster* dengan tujuan ditemukan jarak yang paling dekat dengan setiap data dengan pusat *cluster*. Di bawah ini adalah persamaan jarak *Euclidean*:

$$d(x | j, y_j) = \sqrt{\sum_{i=1}^n (x_i - y_{ij})^2}$$

Keterangan:

- $d(x,y)$: Jarak diantara data pada posisi titik x dan y
- x : Posisi titik data pertama (pusat klaster)
- y : Posisi titik data kedua (data dari n)
- n : Jumlah atribut data

f. Evaluation

Setelah data diolah atau dimodelkan dengan metode *clustering k-means*, kemudian model tersebut dievaluasi. Tahap ini digunakan untuk mengevaluasi kinerja dari berbagai algoritma metode *data mining* untuk mengidentifikasi algoritma yang sesuai untuk diterapkan dalam dataset[13].

g. Knowledge

Knowledge representation didefinisikan sebagai teknik yang menggunakan alat visualisasi untuk mewakili hasil *data mining* seperti menghasilkan dalam bentuk laporan, tabel, aturan klasifikasi, aturan klasifikasi dan lain-lain. *Knowledge process* ini bertujuan untuk memberikan wawasan yang lebih luas perihal masalah kepada pengambil keputusan, selain sebagai pendukung proses optimasi dalam iterasi desain masa depan[14].

RESULTS & DISCUSSION

Result

Data yang digunakan pada penelitian ini didapat dari Open Data Jabar yaitu Jumlah Semua Kasus Tubekulosis Terdaftar di provinsi Jawa Barat. Atribut yang dipakai untuk proses *clustering* yaitu jumlah kasus di kota dan kabupaten pada wilayah provinsi Jawa Barat berdasarkan jenis kelamin pada tahun 2020 dengan jumlah record sebanyak 54. Data tersebut diolah sesuai dengan kebutuhan menggunakan bahasa pemrograman *Python*.

a. Persiapan Data

Tahap ini adalah proses membersihkan, mengubah atau menghapus data mentah sebelum diproses dan dianalisis ke tahap *clustering* dengan tujuan supaya dataset bisa diolah dengan baik. Data yang dipakai yaitu total kasus tuberkulosis di wilayah provinsi Jawa Barat dengan atribut *id*, *kode_provinsi*, *nama_provinsi*, *kode_kabupaten*, *nama_kabupaten_kota*, *jenis_kelamin*, *jumlah_kasus*, *satuan* dan *tahun*. Dari ke-9 atribut tersebut, dilakukan pembersihan data dengan menghapus beberapa atribut yang tidak diperlukan untuk proses *clustering*. Atribut yang digunakan adalah *jenis_kelamin* dan *jumlah_kasus* untuk melihat apakah ada pengaruh jenis kelamin terhadap banyaknya kasus tuberkulosis.

	nama_kabupaten_kota	jenis_kelamin	jumlah_kasus
0	KABUPATEN BOGOR	LAKI-LAKI	5682
1	KABUPATEN BOGOR	PEREMPUAN	4566
2	KABUPATEN SUKABUMI	LAKI-LAKI	1854
3	KABUPATEN SUKABUMI	PEREMPUAN	1592
4	KABUPATEN CIANJUR	LAKI-LAKI	2170

Gambar 1. Informasi Atribut yang digunakan

Kemudian dilakukan pengecekan *missing value* pada atribut yang memiliki data kosong atau *null*. Hasilnya tidak ada data yang bernilai *null*. Lalu dilakukan proses transformasi data yang bertujuan untuk memproses data yang sudah dipilih ke dalam bentuk *mining procedure* melalui cara agregasi data supaya data dapat diolah. Data transformasi yang dilakukan adalah mengubah atribut *categorical* jenis kelamin menjadi

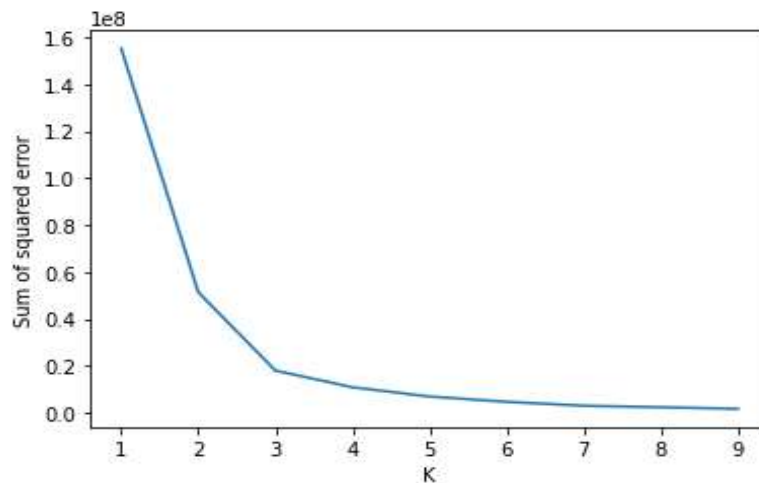
non-categorical untuk meningkatkan akurasi karena algoritma cenderung kesulitan menangani kolom kardinalitas tinggi. Hasil data transformasi dapat dilihat pada Table 1.

Tabel 1. Hasil Transformasi Data

No	jenis_kelamin	jumlah_kasus
0	0	5682
1	1	4566
2	0	1854
3	1	1592
4	0	2170
...
52	1	427
53	0	177
54	1	143

b. Modeling

Selanjutnya dilakukan tahap *modelling data mining*. Hal pertama yang dilakukan yaitu menentukan jumlah *cluster* atau nilai *k* pada data tersebut menggunakan metode *elbow*. Berdasarkan gambar 2, didapat *cluster* optimal yang didapat berjumlah 2 cluster.



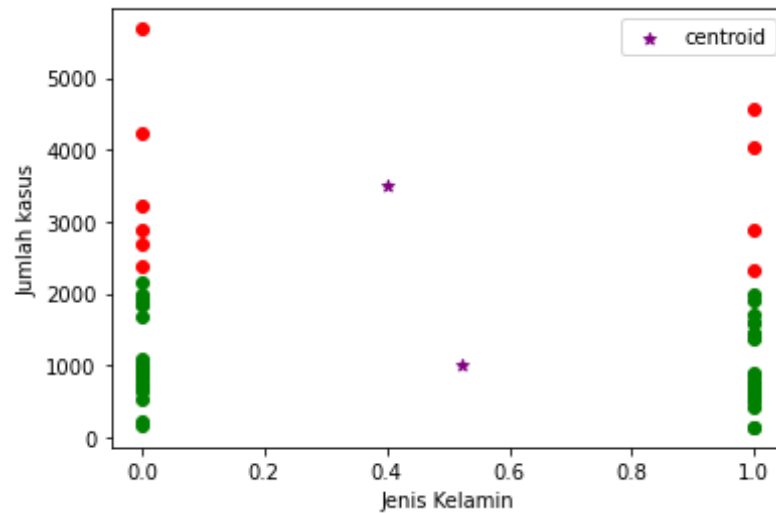
Gambar 2. Total cluster menggunakan metode *Elbow*

Gambar 2 menjelaskan untuk pencarian jumlah cluster, dapat dilihat bahwa dari dataset yang dimiliki tersebut nilai *k* yang paling efektif adalah berjumlah 2, yang berarti ada 2 *cluster* optimal. Selanjutnya adalah menentukan kelompok *cluster* dengan mencari titik pusat atau *centroid* dari data tersebut. Titik *centroid* tersebut bisa dilihat di gambar di bawah ini.

```
km.cluster_centers_
✓ 0.4s
array([[4.00000000e-01, 3.49380000e+03],
       [5.22727273e-01, 1.02050000e+03]])
```

Gambar 3. Titik centroid

Dari hasil pengelompokan tersebut, terdapat 10 data yang masuk kedalam *cluster* 1 yang berarti 10 data tersebut memiliki jumlah kasus tuberkulosis yang tinggi. Lalu sisanya terdapat 44 atribut yang masuk kedalam *cluster* 0 yang berarti terdapat 44 data dengan jumlah kasus rendah. Selanjutnya adalah visualisasi *cluster* 0 dan 1 untuk melihat pengaruh *gender* terhadap jumlah kasus tuberkulosis pada masing-masing daerah pada Gambar 4.



Gambar 4. Pengaruh Gender terhadap jumlah kasus tuberkulosis

Dari gambar diatas, dapat dilihat bahwa jenis kelamin tidak mempengaruhi total kasus tuberkulosis secara signifikan di kabupaten dan kota wilayah provinsi Jawa Barat. Titik berwarna merah adalah daerah kota/kabupaten dengan jumlah kasus tertinggi, sedangkan titik berwarna hijau adalah daerah kota/kabupaten dengan jumlah kasus terendah. Jumlah kasus tertinggi dimiliki oleh Kabupaten Bogor dengan jumlah 5682 kasus pada jenis kelamin laki-laki dan 4566 kasus pada jenis kelamin perempuan pada tahun 2020. Sedangkan jumlah kasus terendah dimiliki oleh Kota Banjar dengan jumlah 177 kasus pada laki-laki dan 143 kasus pada perempuan pada tahun 2020.

Tabel 2. Cluster Tinggi

No	nama_kabupaten_kota	jenis_kelamin	jumlah_kasus
1	KABUPATEN BOGOR	LAKI-LAKI	5682
2	KABUPATEN BOGOR	PEREMPUAN	4566
3	KABUPATEN BANDUNG	LAKI-LAKI	3222
4	KABUPATEN BANDUNG	PEREMPUAN	2894
5	KABUPATEN KARAWANG	LAKI-LAKI	2396
6	KABUPATEN BEKASI	LAKI-LAKI	2693
7	KOTA BANDUNG	LAKI-LAKI	4233
8	KOTA BANDUNG	PEREMPUAN	4029
9	KOTA BEKASI	LAKI-LAKI	2885
10	KOTA BEKASI	PEREMPUAN	2338

Berdasarkan tabel 2 diperlihatkan bahwa kasus tuberkulosis tertinggi adalah

Kabupaten Bogor dengan jumlah kasus 5682 pada jenis kelamin laki-laki dan 4566 pada jenis kelamin perempuan.

Tabel 3. Cluster Rendah

No	nama_kabupaten_kota	jenis_kelamin	jumlah_kasus
1	KABUPATEN SUKABUMI	LAKI-LAKI	2329
2	KABUPATEN SUKABUMI	PEREMPUAN	1845
3	KABUPATEN CIANJUR	LAKI-LAKI	2307
4	KABUPATEN CIANJUR	PEREMPUAN	2202
5	KABUPATEN GARUT	LAKI-LAKI	1667
...
43	KOTA CIMAH	PEREMPUAN	1215
44	KOTA TASIKMALAYA	LAKI-LAKI	705
45	KOTA TASIKMALAYA	PEREMPUAN	653
46	KOTA BANJAR	LAKI-LAKI	400
47	KOTA BANJAR	PEREMPUAN	274

Tabel 3 menunjukkan kasus tuberkulosis terendah. Urutan pertama adalah Kabupaten Sukabumi dengan jumlah kasus 2329 pada jenis kelamin laki-laki dan 1845 pada jenis kelamin perempuan.

c. Evaluation

Tahap evaluasi merupakan tahap untuk melihat kinerja dari model yang sudah dibuat. Pada penelitian ini menggunakan teknik *Silhouette* yang bertujuan untuk menguji kualitas *cluster* dari data yang telah didapatkan sebelumnya pada tahap modelling. Dari pengujian evaluasi ini, didapatkan nilai rata-rata *silhouette* diatas 0.5 yang menandakan kualitas *cluster* cukup baik.

Nilai Rata - Rata Silhouette : 0.5950725534914513

Gambar 5. Rata-rata nilai Silhouette

Pada gambar 5, didapat nilai *silhouette coefficient* sebesar 0.59 dari proses *K-Means Clustering* pada data total kasus tuberkulosis di wilayah Jawa Barat di masing-masing kabupaten/kota. Nilai tersebut termasuk pada kriteria *Medium Structure* pada urutan kriteria dari *silhouette coefficient*. Nilai tersebut sudah dapat dikatakan baik jika nilainya > 0.5 . Jadi pengujian kualitas *cluster* pada jumlah kasus tuberkulosis di Jawa Barat mempunyai kualitas cukup baik.

d. Knowledge

Berdasarkan penelitian yang sudah dilakukan menggunakan Metode *K-Means Clustering* pada data total Kasus Tuberkulosis di wilayah Jawa Barat pada tahun 2020 menghasilkan data bahwa *gender* atau jenis kelamin tidak mempengaruhi banyaknya jumlah kasus tuberkulosis. Pada penelitian ini juga menghasilkan 2 *cluster* terhadap tinggi dan rendahnya jumlah kasus di provinsi jawa barat. Tabel 4 adalah hasil *cluster* pada pengolahan data diatas menggunakan *k-means clustering*.

Tabel 4. Hasil Data dari pengujian *Cluster* total kasus Tuberkulosis di wilayah Provinsi Jawa Barat Berdasarkan Jenis Kelamin

No. Cluster	Kabupaten/Kota
0	Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Subang, Kabupaten Purwakarta, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar, Kabupaten Bekasi (Perempuan)
1	Kabupaten Bogor, Kabupaten Bandung, Kabupaten Karawang (Laki-laki), Kabupaten Bekasi (Laki-laki), Kota Bandung, Kota Bekasi

Dari tabel tersebut, *cluster* 0 adalah kelompok dengan jumlah kasus tuberkulosis rendah, sedangkan *cluster* 1 adalah kelompok dengan jumlah kasus tuberkulosis tinggi.

DISCUSSION

Berdasarkan penelitian yang dilakukan oleh peneliti pada data total masalah tuberkulosis di wilayah provinsi Jawa Barat, dapat ditarik kesimpulan bahwa jenis kelamin atau *gender* tidak mempengaruhi jumlah kasus tuberkulosis, karena hasil *cluster* menunjukkan tidak ada perbedaan signifikan dalam total kasus antara jenis kelamin laki-laki dan jenis kelamin perempuan. Hasil penelitian juga menghasilkan 2 *cluster* yaitu *cluster* dengan jumlah kasus tinggi dan rendah. Jumlah kasus dengan *cluster* tinggi terdapat 6 daerah, sedangkan *cluster* terendah terdapat 21 daerah. Evaluasi menggunakan *silhouette coefficient* mendapatkan nilai sebesar 0.59 yang berarti hasil *cluster* dari penelitian ini cukup baik..

CONCLUSION

Hasil dari penelitian ini diharapkan menjadi perhatian pemerintah kabupaten/kota di provinsi Jawa Barat dalam menekan jumlah kasus di daerahnya masing-masing, khususnya pada daerah dengan *cluster* tinggi. Riset ini dapat pula dipakai sebagai data referensi atau komparasi bagi riset berbeda yang hendak mengambil topik yang sama yang bertujuan mengetahui hasil dari total kasus tuberkulosis di tahun-tahun kedepan.

CONFLICT OF INTEREST

Mengenai penelitian, kepenulisan, dan publikasi makalah ini, penulis melaporkan tidak ada potensi konflik kepentingan.

ACKNOWLEDGEMENT

Kami ingin mengucapkan terima kasih kepada pihak yang sudah membantu dalam penulisan artikel ini, khususnya dosen pembimbing kami yaitu Tesa Nur Padilah, S.Si., M.Sc.

REFERENCES

- Dinas Kesehatan, "Jumlah Kasus Tuberkulosis Berdasarkan Jenis Kelamin di Jawa Barat," *opendata.jabarprov.go.id*, 2021. <https://opendata.jabarprov.go.id/id/dataset/jumlah-kasus-tuberkulosis-berdasarkan-jenis-kelamin-di-jawa-barat> (accessed Dec. 03, 2021).
- E. Wahyudi, "Hindari Corona, Tokopedia Instruksikan Karyawan Bekerja di Rumah," *bisnis.tempo.co*, 2020. <https://bisnis.tempo.co/read/1320278/hindari-corona-tokopedia-instruksikan-karyawan-bekerja-di-rumah> (accessed Mar. 20, 2020).
- G. Yuan, P. Sun, J. Zhao, D. Li, and C. Wang, "A review of moving object trajectory clustering algorithms," *Artif. Intell. Rev.*, vol. 47, no. 1, pp. 123–144, 2017, doi: 10.1007/s10462-016-9477-7.
- H. Ltifi, C. Kolski, and M. Ben Ayed, "Combination of cognitive and HCI modeling for the design of KDD-based DSS used in dynamic situations," *Decis. Support Syst.*, vol. 78, pp. 51–64, 2015, doi: 10.1016/j.dss.2015.07.003.
- M. Fauzi and Y. Yudi, "Penerapan Algoritma K-Means Clustering Untuk Mendeteksi Penyebaran Penyakit TBC (Studi Kasus: Di Kabupaten Deli Serdang)," *JTIK (Jurnal Tek. Inform. Kaputama)*, vol. 1, no. 2, pp. 1–7, 2017, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/download/41/38>.
- M. C. Untoro, L. Anggraini, M. Andini, H. Retnosari, and M. A. Nasrulloh, "Penerapan metode k-means clustering data COVID-19 di Provinsi Jakarta," *Teknologi*, vol. 11, no. 2, pp. 59–68, 2021, doi: 10.26594/teknologi.v11i2.2323
- N. A. Menzies *et al.*, "Lifetime burden of disease due to incident tuberculosis: a global reappraisal including post-tuberculosis sequelae," *Lancet Glob. Heal.*, vol. 9, no. 12, pp. e1679–e1687, 2021, doi: 10.1016/S2214-109X(21)00367-3.
- S. Umadevi and K. S. J. Marseline, "A survey on data mining classification algorithms," *Proc. IEEE Int. Conf. Signal Process. Commun. ICSPC 2017*, vol. 2018-January, no. July, pp. 264–268, 2018, doi: 10.1109/CSPC.2017.8305851.
- S. Hussain, L. J. Muhammad, F. S. Ishaq, A. Yakubu, and I. A. Mohammed, *Performance evaluation of various data mining algorithms on road traffic accident dataset*, vol. 106. Springer Singapore, 2019.
- S. Bandaru, A. H. C. Ng, and K. Deb, *Data mining methods for knowledge discovery in multi-objective optimization: Part A - Survey*, vol. 70. Elsevier Ltd, 2017.
- T. Siswantining, N. P. C. Dewi Purwandani, M. H. Dewi Susilowati, and A. Wibowo, "Geoinformatics of tuberculosis (TB) disease in Jakarta city Indonesia," *Int. J. GEOMATE*, vol. 19, no. 72, pp. 35–42, 2020, doi: 10.21660/2020.72.5599.
- 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.
- R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda)," *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.
- V. Shunmughavel, "Weather Forecast Warning System using Big Data and Naive Bayes Classification Algorithm," vol. IX, no. Viii, pp. 363–370, 2020.