

Penerapan K-Means *Clustering* pada Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Kabupaten Karawang

Isy Karima Fauzia¹, Budi Arif Dermawan², Tesa Nur Padilah³

Universitas Singaperbangsa Karawang

e-mail: ¹isy.16121@student.unsika.ac.id, ²budi.arif@staff.unsika.ac.id,

³tesa.nurpadilah@staff.unsika.ac.id

Diajukan: 23 Oktober 2020; Direvisi: 22 November 2020; Diterima:

Abstrak

Infeksi Saluran Pernapasan Akut (ISPA) adalah penyakit yang menyerang pernapasan bawah dan pernapasan atas dapat mengakibatkan kematian. Menurut Badan Pusat Statistik Kabupaten Karawang pada tahun 2017 melaporkan 173.953 kasus ISPA dan meningkat menjadi 175.891 pada tahun 2018. Belum adanya perhatian khusus pada daerah tersebut yang menjadikan penyakit ini terus meningkat. Pada penelitian ini adalah clustering penyakit ISPA menggunakan algoritma K-Means dengan Cross-Industry Standard Process for Data Mining (CRISP-DM). Di mana pada penelitian sebelumnya menyimpulkan bahwa teknik clustering yang paling optimal adalah metode K-Means karena hasilnya lebih akurat dalam pengelompokan data dan tidak menggunakan tahap CRISP-DM. Untuk itu diperlukannya clustering daerah penyebaran ISPA yang terbagi menjadi tiga cluster yaitu rendah, sedang, dan tinggi sehingga dapat membantu dalam pengambilan kebijakan terhadap penyakit ISPA oleh pemerintah Kabupaten Karawang. Hasilnya adalah tahun 2017 terdapat 30 anggota cluster rendah, 9 anggota cluster sedang, dan 11 anggota cluster tinggi. Tahun 2018 terdapat 33 anggota cluster rendah, 5 anggota cluster sedang, dan 12 anggota cluster tinggi. Tahun 2019 terdapat 20 anggota cluster rendah, 25 anggota cluster sedang, dan 5 anggota cluster tinggi. Evaluasi algoritma K-Means menggunakan SSE pada tahun 2017 yaitu 232.6133, 2018 yaitu 207.8584, dan tahun 2019 yaitu 260.3935.

Kata kunci: Data mining, Clustering, Infeksi Saluran Pernapasan Akut, K-Means, Sum of Square Error.

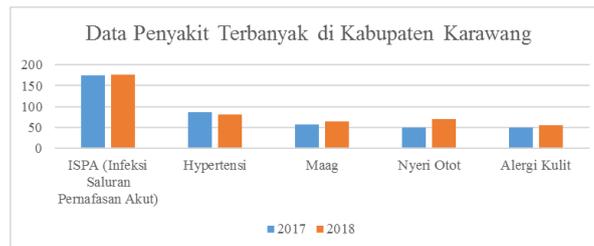
Abstract

Acute Respiratory Infection (ARI) is a disease that attacks the lower respiratory and upper respiratory which can result in death. The Karawang Central Statistics Agency in 2017 reported 173,953 cases of ARI and increased to 175,891 in 2018. The lack of special attention in the area of ISPA distribution makes this disease continue to increase. In this research clustering disease ARI using the K-Means algorithm with Cross-Industry Standard Process for Data Mining (CRISP-DM). Where the previous research concluded that the most optimal clustering technique was the K-Means method because the results were more accurate in grouping data and did not use the CRISP-DM stage. The aim is to classify areas of ISPA distribution into three clusters, namely low, medium, and high so that they can help in making decisions about ARI by the Karawang regency government. The result is that in 2017 there were 30 low cluster members, 9 moderate cluster members, and 11 high cluster members. In 2018 there are 33 low cluster members, 5 medium cluster members, and 12 high cluster members. In 2019 there are 20 low cluster members, 25 medium cluster members, and 5 high cluster members. Evaluation of K-Means algorithm using SSE in 2017 is 232.6133, 2018 is 207.8584, and 2019 is 260.3935.

Keywords: Acute Respiratory Infection, Data mining, Clustering, K-Means, Sum of Square Error.

1. Pendahuluan

Infeksi Saluran Pernapasan Akut (ISPA) adalah penyakit yang sering terjadi pada masyarakat. ISPA bisa menyerang semua umur terutama pada anak-anak baik pernapasan bawah atau atas, dan dapat menyebabkan berbagai penyakit dari infeksi ringan sampai parah yang dapat mengakibatkan kematian [1]. Infeksi Saluran Pernapasan Akut (ISPA) yang menyerang bagian bawah pernapasan disebut pneumonia [2]. Pneumonia adalah infeksi akut yang mengenai jaringan paru-paru yang disebabkan oleh berbagai bakteri, virus, dan jamur [3].



Gambar 1. Data penyakit terbanyak di Kabupaten Karawang.

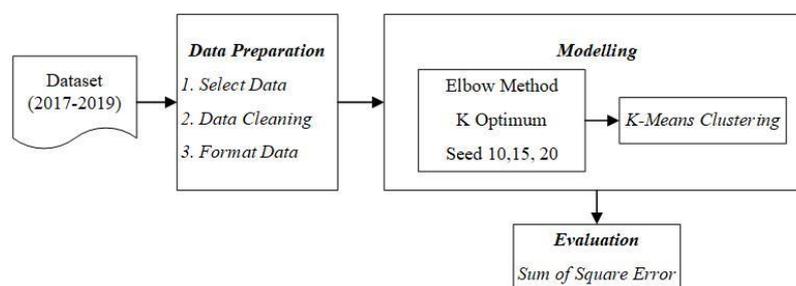
Badan Pusat Statistik Kabupaten Karawang melaporkan bahwa pada tahun 2017 ditemukan kasus ISPA sebanyak 173.953 [4] dan di tahun 2018 ditemukan kasus ISPA sebanyak 175.891 [5]. Gambar 1 menunjukkan bahwa selama 2 tahun berturut-turut penyakit ISPA menjadi penyakit peringkat pertama di Kabupaten Karawang. Situasi saat ini perlakuan Dinas Kesehatan Karawang terhadap semua puskesmas sama, tidak ada perhatian khusus pada salah satu puskesmas. Oleh karena itu sebaiknya pemerintah Karawang perlu melakukan penanganan yang efektif terhadap penyakit ISPA. Dalam membantu penanganan yang efektif terhadap penyakit ISPA diperlukan suatu metode untuk menemukan daerah penyebaran penyakit ISPA di Kabupaten Karawang. Tujuannya agar pemerintah Karawang dapat memberi perhatian khusus, penanganan yang efektif, dan dapat membantu dalam pengambilan kebijakan terhadap penyakit ISPA di Kabupaten Karawang misalnya puskesmas mana saja yang memerlukan perhatian khusus atau perhatian lebih.

Beberapa penelitian yang telah dilakukan mengenai pengelompokan menggunakan algoritma K-Means. Penelitian analisis pada penyakit menular manusia di Kabupaten Majalengka disimpulkan bahwa K-Means merupakan salah satu metode *clustering*, alasan menggunakan algoritma K-Means mempunyai ketelitian yang optimal terhadap ukuran objek sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar [6]. Penelitian lainnya menyimpulkan bahwa teknik *clustering* yang paling optimal adalah metode K-Means karena hasilnya lebih akurat dalam pengelompokan data [7]. Penelitian menghasilkan pengetahuan baru daerah yang banyak terkena penyakit ISPA dengan tahap *Knowledge Discovery in Database* (KDD) menggunakan *tool* RapidMiner dan disimpulkan bahwa algoritma K-Means mendapatkan hasil yang optimal untuk pengklasteran [8].

Dari permasalahan di atas maka akan dilakukan penelitian *clustering* penyakit ISPA menggunakan algoritma K-Means dengan tahap *Cross-Industry Standard Process for Data mining* (CRISP-DM). *Tool* yang digunakan adalah RStudio dengan menghitung akurasi nilai *error* menggunakan metode SSE (*Sum of Square Error*). Data yang digunakan adalah data penyakit ISPA dari tahun 2017-2019 yang didapat dari Dinas Kesehatan Kabupaten Karawang dengan tipe data numerik. Penelitian ini dilakukan dengan harapan untuk mempermudah Dinas Kesehatan dalam menemukan daerah penyebaran penyakit ISPA.

2. Metode Penelitian

Alur penelitian diilustrasikan pada Gambar 2, yang terdiri dari 3 bagian utama yaitu *data preparation*, *modelling*, dan *evaluation*.



Gambar 2. Alur penelitian.

2.1. Dataset

Dataset yang didapatkan terdiri dari 11 variabel. Berikut adalah keterangan *dataset* tersebut.

Tabel 1. Keterangan *dataset*.

Tahun	Atribut	Sumber
2017	Nomor, Puskesmas, Jumlah Penduduk, Pneumonia umur (<1 th), Pneumonia umur (1 th -<5 th), Pneumonia umur (≥5 th), Bukan Pneumonia umur (<1 th), Bukan Pneumonia umur (1 th -<5 th), Bukan Pneumonia umur (≥5 th), Total Pneumonia, Total Bukan Pneumonia	Dinas Kesehatan Kabupaten Karawang
2018	Nomor, Puskesmas, Jumlah Penduduk, Pneumonia umur (<1 th), Pneumonia umur (1 th -<5 th), Pneumonia umur (≥5 th), Bukan Pneumonia umur (<1 th), Bukan Pneumonia umur (1 th -<5 th), Bukan Pneumonia umur (≥5 th), Total Pneumonia, Total Bukan Pneumonia	Dinas Kesehatan Kabupaten Karawang
2019	Nomor, Puskesmas, Jumlah Penduduk, Pneumonia umur (<1 th), Pneumonia umur (1 th -<5 th), Pneumonia umur (≥5 th), Bukan Pneumonia umur (<1 th), Bukan Pneumonia umur (1 th -<5 th), Bukan Pneumonia umur (≥5 th), Total Pneumonia, Total Bukan Pneumonia	Dinas Kesehatan Kabupaten Karawang

2.2. Imputasi Mean

Imputasi dengan *mean* merupakan salah satu imputasi yang sering digunakan. *Mean* adalah mengambil nilai rata-rata dari semua nilai data yang diketahui. Metode ini hanya bisa digunakan untuk data yang berbentuk numerik saja. Untuk mendapatkan nilai *mean* maka digunakan dengan rumus sebagai berikut [9].

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \tag{1}$$

\bar{x} = Nilai imputasi baris

x_i = Nilai baris

n = Banyaknya data pada atribut

2.3. Algoritma K-Means

K-Means merupakan salah satu algoritma *clustering* yang dapat digunakan untuk mengelompokkan suatu data, data yang memiliki kemiripan yang tinggi dikelompokkan dalam satu *cluster* sedangkan data yang memiliki karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang berbeda [7]. K-Means termasuk ke dalam salah satu metode *clustering* yang berjenis *non-hierarchical*. Berikut adalah proses langkah-langkah dasar melakukan *clustering* dengan menggunakan metode K-Means [10].

1. Menentukan jumlah *cluster* yang ingin dibentuk dan tetapkan pusat *cluster* k.
2. Pusat awal *cluster* (*centroid*) dipilih secara acak.
3. Menghitungnya menggunakan jarak setiap data *input* terdapat masih-masih *cluster* awal dengan rumus jarak Euclidean (Euclidean Distance) sampai menemukan jarak paling dekat dari setiap data dengan *cluster* awal. Berikut rumus persamaan Euclidean Distance:

$$D(x_i, \pi_i) = \sqrt{\sum_{i=1}^n (x_i - \pi_i)^2} \tag{2}$$

Keterangan:

$D(x_i, \pi_i)$ = Jarak antara *clustering* x dengan pusat *cluster* μ pada data ke-i

x_i = Bobot data ke-i pada *cluster* yang ingin dicari jaraknya

μ_i = Bobot data ke-i pada pusat *cluster*

n = Jumlah data

4. Kelompokkan data berdasarkan jarak yang terdekat antara data dengan *cluster* awal.
5. Tentukan nilai *cluster* awal yang baru dengan cara hitung rata-rata dari *cluster* yang bersangkutan dengan rumus sebagai berikut:

$$C_k = \frac{1}{n_k} \sum_{i=1}^n d_i \tag{3}$$

Keterangan:

n_k = total data dalam *cluster* k

d_i = total pada masing-masing *cluster*

6. Ulangi langkah tiga sampai dengan lima sampai setiap anggota *cluster* tidak ada lagi yang berubah.

2.4. Sum of Square Error (SSE)

Validasi data dapat diperoleh dengan menghitung nilai SSE (*Sum of Square Error*) dari masing-masing nilai *cluster*. Semakin besar jumlah *cluster* k maka nilai SSE akan semakin kecil [10]. Untuk menghitung SSE menggunakan rumus sebagai berikut:

$$SSE = \sum_{k=1}^{K=n} \sum_{xi \in S_k} \|X_i - C_k\| \tag{2}$$

Keterangan:

- K = total *cluster*
- X_i = data ke- i
- C_k = awal *cluster*
- n = total data

3. Hasil dan Pembahasan

3.1. Data preparation

3.1.1. Select Data

Data diambil tentang penyakit ISPA di Kabupaten Karawang dan disimpan dalam bentuk *file* Microsoft Excel. Lalu data dibersihkan dengan cara menghapus atribut-atribut yang tidak diperlukan pada penelitian yang dilakukan. Data yang dipilih adalah data yang mengenai penyakit ISPA antara tahun 2017-2019. Atribut yang digunakan di antaranya nama puskesmas, jumlah penduduk, kelompok pneumonia, bukan pneumonia berdasarkan usia (<1th), usia (1th-<5th), usia (≥5th), total pneumonia, dan total bukan pneumonia.

3.1.2. Data Cleaning

Terdapat beberapa data yang ketika di-*input* kembali dalam *file* Excel memiliki hasil penjumlahan berbeda dengan data *hardcopy*. Oleh karena itu untuk *dataset final* menggunakan hasil penjumlahan yang dihitung pada *file* Excel. Pada tahap ini dilakukan proses imputasi menggunakan teknik imputasi *mean*.

Tabel 2. Hasil *data cleaning dataset* tahun 2017.

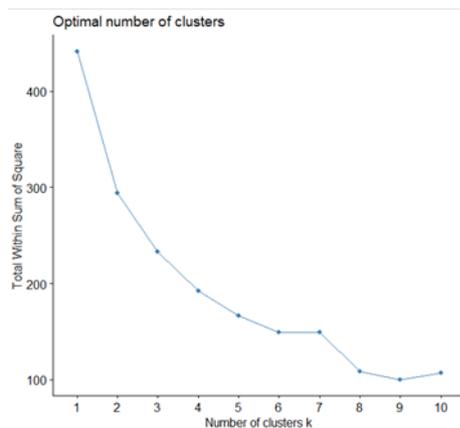
Puskesmas	Jumlah Penduduk	Pneumonia			Total Pneumonia	Bukan Pneumonia			Total Bukan Pneumonia
		(<1 th)	(1 th -<5 th)	(≥5 th)		(<1 th)	(1 th -<5 th)	(≥5 th)	
Adiarsa	43888	19	60	3	82	427	682	1185	2294
Anggadita	23561	20.26	46	44	90	147.2	186	321	507
Balongsari	17663	3	16	17.96	19	21	101	490	612
Batujaya	81167	47	91	17.96	138	65	119	183	367
Bayur Lor	26437	20.26	26.46	17.96	64.68	147.2	236.8	557.5	941.4
Ciampel	39113	4	10	5	19	87	148	578	813
Cibuaya	59649	117	127	116	360	89	142	286	517
Cicinde	31165	24	52	24	100	282	512	1568	2362
Cikampek	111415	20.26	26.46	17.96	64.68	147.2	236.8	557.5	941.4
Utara	36991	20.26	26.46	17.96	64.68	147.2	236.8	557.5	941.4
Cilamaya	56182	20.26	26.46	17.96	64.68	147.2	236.8	557.5	941.4
...
Wadas	49361	13	21	17.96	34	418	488	557.5	906
Wanakerta	49042	5	6	17.96	11	28	121	946	1095

3.1.3. Format Data

Tahap ini memproduksi *dataset* akhir yang siap ditambah atau diolah dalam *tool* pemodelan *data mining*. Data yang diperoleh dari Dinas Kesehatan Kabupaten Karawang dengan *hardcopy*, berupa *file* Microsoft Excel dengan format .xlsx diubah menjadi format CSV (*Comma Separated Values*). Hal ini untuk mempermudah proses modeling yang akan dilakukan di RStudio.

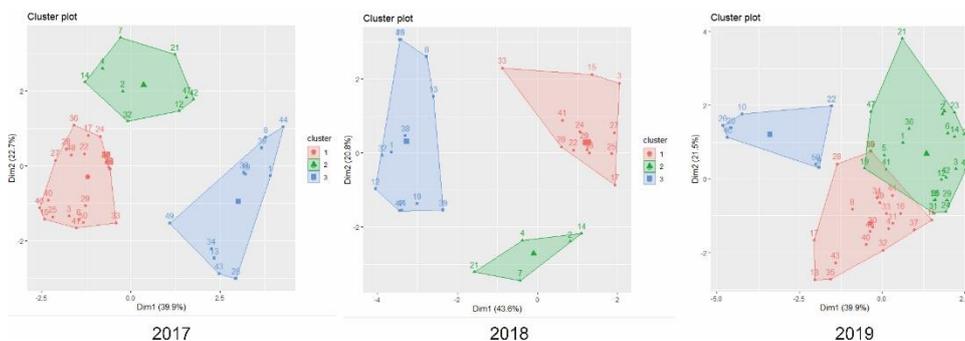
3.2. Modelling

Dataset yang telah melalui tahap *data preparation* selanjutnya dilakukan tahap *modelling*. Hal pertama yang dilakukan yaitu menentukan jumlah *cluster* dengan menggunakan metode Elbow. Berdasarkan Gambar 3, *cluster* optimal yang didapat yaitu 3 *cluster*.



Gambar 3. Grafik *cluster* optimal dengan metode Elbow.

Setelah didapatkan jumlah *cluster* optimal, selanjutnya adalah tahap pembangunan model dengan menggunakan algoritma K-Means dengan kombinasi *seed* yaitu 10, 15, dan 20. Hasil terbaik dari yaitu tahun 2017 dengan *seed* 15, tahun 2018 dengan *seed* 10, dan tahun 2019 dengan *seed* 15. Berikut adalah plot dari tahun masing-masing.



Gambar 4. Plot *cluster* per tahun.

Pada Gambar 4 dapat dilihat bahwa titik tengah merupakan nilai *centroid* dari masing-masing *cluster*. Tahun 2017, *cluster* satu merupakan *cluster* dengan jumlah anggota terbanyak yaitu 31 anggota, *cluster* dua memiliki 9 anggota, dan *cluster* tiga memiliki 11 anggota. Sedangkan pada tahun 2018, *cluster* satu memiliki 33 anggota, *cluster* dua memiliki 5 anggota, dan *cluster* tiga memiliki 12 anggota. Lalu tahun 2019, *cluster* satu memiliki 25 anggota, *cluster* dua memiliki 5 anggota, dan *cluster* tiga memiliki 19 anggota. Berikut adalah rata-rata total kasus ISPA per tahun.

Tabel 3. Kategori ISPA per *cluster*.

<i>Cluster</i>	2017	Kategori	2018	Kategori	2019	Kategori
<i>Cluster</i> 1	1509.37	Rendah	1263.64	Rendah	7405.76	Sedang
<i>Cluster</i> 2	2384.15	Sedang	1345.98	Sedang	12260.15	Tinggi
<i>Cluster</i> 3	5997.17	Tinggi	4812.92	Tinggi	2868.96	Rendah

Pada Tabel 3 dapat dilihat bahwa tidak ada perbedaan antara tahun 2017 dan 2018. Hanya terdapat perbedaan pada tahun 2019 yaitu *cluster* satu dari rendah menjadi sedang, *cluster* dua dari sedang menjadi tinggi, dan *cluster* tiga dari tinggi menjadi rendah. Berikut adalah anggota dari tiap *cluster*.

Tabel 4. Anggota per *cluster*.

Tahun	Cluster 1	Cluster 2	Cluster 3
2017	Balongsari, Bayur Lor, Ciampel, Cikampek, Cikampek Utara, Cilamaya, Jatisari, Jayakarta, Jomin, Kalangsari, Karawang, Kertamukti, Kota Baru, Kutamukti, Kutawaluya, Lemah Duhur, Loji, Majalaya, Medang Asem, Nagasari, Pacing, Pangkalan, Pedes, Plawad, Purwasari, Sukatani, Sungai Buntu, Tempuran, Tirtajaya, Tunggak Jati, Wanakerta.	Anggadita, Batujaya, Cibuyaya, Curug, Jatisari, Klari, Pakisjaya, Tanjungpura, Tirtamulya	Adiarsa, Cicinde, Gempol, Karawang Kulon, Lemahabang, Pasirukem, Rawamerta, Rengsdengklok, Telagasari, Telukjambe, Wadas
2018	Balongsari, Bayur Lor, Ciampel, Cikampek, Cikampek Utara, Cilamaya, Jayakarta, Jomin, Kalangsari, Karawang, Kertamukti, Kota Baru, Kutamukti, Kutawaluya, Lemah Duhur, Loji, Majalaya, Medang Asem, Nagasari, Pacing, Pangkalan, Pasirukem, Pedes, Plawad, Purwasari, Sukatani, Sungai Buntu, Tanjungpura, Tempuran, Tirtajaya, Tunggak Jati, Wada, Wanakerta	Anggadita, Batujaya, Cibuyaya, Jatisarim, Klari	Adiarsa, Cicinde, Curug, Gempol, Karawang Kulon, Lemahabang, Pakisjaya, Rawamerta, Rengsdengklok, Telagasari, Teluk Jambe, Tirtamulya
2019	Batujaya, Bayur Lor, Cicinde, Cikampek, Cilamaya, Gempol, Jomin, Kalangsari, Karawang Kulon, Majalaya, Nagasari, Pakisjaya, Pangkalan, Pasirukem, Pedes, Purwasari, Rawamerta, Rengsdengklok, Sukatani, Sungai Buntu, Telagasari, Teluk Jambe, Wadas, Wanakerta	Cikampek Utara, Klari, Kota Baru, Lemahabang, Tirtajaya	Adiarsa, Anggadita, Balongsari, Cibuyaya, Ciampel, Curug, Jatisari, Jayakarta, Karawang, Kertamukti, Kutawaluya, Lemah Duhur, Loji, Medang Asem, Pacing, Plawad, Tanjungpura, Tirtamulya, Tunggak Jati

3.3. Evaluation

Setelah dilakukan tahap *modelling*, selanjutnya adalah mengevaluasi model menggunakan *Sum of Square Error* (SSE). Berikut adalah hasil dari SSE.

Tabel 5. Nilai *Sum of Square Error*.

Tahun	S=10	S=15	S=20
2017	232.6891	232.6133	252.6942
2018	207.8584	207.8584	212.4581
2019	260.6179	260.3935	260.6179

Berdasarkan Tabel 5 dapat dilihat bahwa nilai SSE terbaik tahun 2017 yaitu dengan *seed* 15, tahun 2018 dengan *seed* 10, dan tahun 2019 dengan *seed* 15.

3.4. Pembahasan

Dari hasil penelitian didapatkan bahwa tahun 2017, *cluster* 1 merupakan *cluster* dengan jumlah anggota terbanyak yaitu 31 anggota dikategorikan sebagai rendah dengan peringkat pertama total 1029 kasus, *cluster* 2 memiliki 9 anggota dikategorikan sebagai sedang dengan peringkat pertama total 1197 kasus, dan *cluster* 3 memiliki 11 anggota dikategorikan sebagai tinggi dengan peringkat pertama total 2774 kasus. Sedangkan pada tahun 2018, *cluster* 1 memiliki 33 anggota dikategorikan sebagai rendah dengan peringkat pertama total 769 kasus, *cluster* 2 memiliki 5 anggota dikategorikan sebagai tinggi dengan peringkat pertama total 169 kasus, dan *cluster* 3 memiliki 12 anggota dikategorikan sebagai sedang dengan peringkat pertama total 713 kasus. Lalu tahun 2019, *cluster* 1 memiliki 25 anggota dikategorikan sebagai sedang dengan peringkat pertama total 6541 kasus, *cluster* 2 memiliki 5 anggota dikategorikan sebagai tinggi dengan peringkat pertama total 2628 kasus, dan *cluster* 3 memiliki 19 anggota dikategorikan sebagai rendah dengan peringkat pertama total 1514 kasus. Salah satu penyebabnya adalah polusi udara dan banyaknya pabrik-pabrik yang menimbulkan polusi.

4. Kesimpulan

Data yang dihasilkan dalam penerapan algoritma K-Means akan mendekati ketepatan dalam pengelompokan daerah yang mempunyai kasus penyakit ISPA di Kabupaten Karawang sehingga mendapatkan perhatian khusus agar dapat mengurangi jumlah kasus ISPA di daerah tersebut. Saran untuk penelitian selanjutnya menambah variabel lain yang berhubungan dengan penyakit ISPA misalnya tingkat kepadatan penduduk, faktor lingkungan, dan sosial, dan penelitian selanjutnya bisa menggunakan algoritma lain.

Daftar Pustaka

[1] D. Maharani, F. F. Yani, and Y. Lestari, "Profil Balita Penderita Infeksi Saluran Nafas Akut Atas

Penerapan K-Means Clustering pada Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Kabupaten Karawang (Isy Karima Fauzia)

-
- di Poliklinik Anak RSUP DR. M. Djamil Padang Tahun 2012-2013,” *J. Kesehat. Andalas*, vol. 6, no. 1, pp. 152–157, 2017.
- [2] S. Hayati, “Gambaran Faktor Penyebab Infeksi Saluran Pernafasan Akut (Ispa) pada Balita di Puskesmas Pasirkaliki Kota Bandung,” *J. Keperawatan BSI*, vol. 2, no. 1, 2014.
- [3] K. K. R. Indonesia, “Profil Kesehatan Indonesia 2016,” *Terdapat di www.depkes.go.id/.../profil-kesehatan-indonesia/Profil-Kesehatan-Indonesia-2016.pdf*. Diakses pada, vol. 20, 2017.
- [4] B. P. S. Karawang, “Kabupaten Karawang dalam Angka 2018 (Karawang Regency in Figure).” Karawang, BPS Karawang, 2018.
- [5] B. P. S. Karawang, “Kabupaten Karawang dalam Angka 2019 (Karawang Regency in Figure).” Karawang, BPS Karawang, p. 2019.
- [6] A. Bastian, H. Sujadi, and P. A. Sukmana, “Rancang bangun aplikasi penilaian ujian essay dengan menggunakan algoritma nazief & andriani dan metode cosine similarity,” *INFOTECH J.*, vol. 4, no. 2, pp. 62–68, 2018.
- [7] N. Butarbutar, A. P. Windarto, D. Hartama, and S. Solikhun, “Komparasi Kinerja Algoritma Fuzzy C-Means Dan K-Means Dalam Pengelompokan Data Siswa Berdasarkan Prestasi Nilai Akademik Siswa,” *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 1, no. 1, pp. 46–55, 2017.
- [8] T. A. Sari, M. Megawaty, and A. Putra, “Pengelompokan Penyebaran Penyakit Ispa Di Wilayah Kota Sekayu Menggunakan Algoritma K-Means Clustering (Studi Kasus: Rsud Sekayu),” in *Bina Darma Conference on Computer Science (BDCCS)*, 2019, vol. 1, no. 1, pp. 174–184.
- [9] A. Izzah and N. Hayatin, “Imputasi Missing data Menggunakan Algoritma Pengelompokan Data K-Harmonic Means,” in *Seminar Nasional Matematika dan Aplikasinya*, 2013.
- [10] A. T. Rahman, “Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama),” *ITSMART J. Teknol. dan Inf.*, vol. 6, no. 1, pp. 24–31, 2017.