

PENGELOMPOKAN PENYAKIT BERDASARKAN LINGKUNGAN DENGAN ALGORITMA K-MEANS PADA PUSKESMAS SUNGAI TARAB 2

Dian Permata Sari

Program Studi Sistem Informasi, STMIK Jayanusa, Jl. Olo Ladang No.1 Padang Tlp (0751) 28984
email: Langitbiru621@gmail.com

Abstract

Sungai Tarab 2 Community Health Center serves at least 60 patients every day. This data continues to grow along with the increasing number of patients who seek treatment so that the large amount of data cannot be studied at all and ends up being archives. Based on the background of the problem above, the writer wants to process the data by classifying diseases based on the environment with those that are not based on the environment by applying the K-Means method of data mining techniques. K-Means is a clustering algorithm that is used to group data into several groups with a partition system. So that later it can help the puskesmas in providing education to the community about the importance of environmental hygiene and diseases that arise due to an unsanitary environment. Based on the results of testing the k-means algorithm on 15 data samples, it is obtained cluster_0 or diseases based on the environment as many as 10 items while the diseases included in the environment are malaria, dengue fever and dermatitis. Meanwhile, cluster_1 or non-environmental diseases have 5 items and the diseases included in this group are vertigo and bronchitis. Based on the data above, it can be concluded that diseases based on the environment are still suffered by many people and education is needed from the puskesmas so that people can avoid these diseases.

Keywords : data mining, k-means, clustering, disease, environment

Abstrak

Puskesmas sungai tarab 2 setidaknya melayani sekitar 60 pasien setiap harinya. Data ini terus bertambah seiring dengan bertambahnya pasien yang berobat sehingga jumlah data yang begitu banyak tidak dapat dipelajari semuanya dan berakhir menjadi arsip saja. Berdasarkan latar belakang masalah diatas penulis ingin mengolah data tersebut dengan mengelompokkan penyakit berdasarkan lingkungan dengan yang bukan berdasarkan lingkungan menerapkan teknik data mining metode K-Means. K-Means adalah algoritma Clustering yang digunakan untuk mengelompokkan data kedalam beberapa kelompok dengan sistem partisi. Sehingga nantinya dapat membantu pihak puskesmas dalam memberikan penyuluhan kepada masyarakat akan pentingnya kebersihan lingkungan serta penyakit-penyakit yang timbul akibat lingkungan kurang bersih. Berdasarkan hasil pengujian algoritma k-means terhadap 15 sampel data maka diperoleh cluster_0 atau penyakit berdasarkan lingkungan sebanyak 10 items adapaun penyakitnya yang termasuk berdasarkan lingkungan adalah malaria, dbd dan dermatitis. Sedangkan cluster_1 atau penyakit bukan berdasarkan lingkungan sebanyak 5 items dan penyakit yang termasuk dalam kelompok ini adalah vertigo dan brongkitis. Berdasarkan data di atas dapat disimpulkan bahwa penyakit berdasarkan lingkungan masih banyak diderita masyarakat dan diperlukan edukasi dari pihak puskesmas agar masyarakat dapat terhindar dari penyakit tersebut.

Kata kunci : data Mining, k-means, clustering, penyakit, lingkungan

1. PENDAHULUAN

Kesehatan merupakan hak asasi manusia. Setiap manusia memiliki hak yang sama untuk memperoleh layanan kesehatan yang baik.

Di Indonesia terdapat puskesmas yang merupakan suatu lembaga milik pemerintah dan berfungsi untuk melayani setiap masyarakat yang membutuhkan pengobatan. Puskesmas

mampu menjangkau masyarakat sampai ke pelosok negeri.

Puskesmas merupakan salah satu bentuk upaya pemerintah untuk memberikan pelayanan kesehatan kepada masyarakat. (Irmawati, 2017). Puskesmas sungai tarab 2 merupakan salah satu puskesmas yang terletak di kabupaten Tanah Datar Provinsi Sumatera barat. Sebagai instansi pemerintah dibawah dinas kesehatan puskesmas sungai tarab melakukan fungsinya yaitu memberikan pelayanan kepada masyarakat. Setiap hari nya puskesmas ini beroperasi mulai dari pukul 8.00 WIB sampai puku 15.00 dan melayani lebih kurang 60 pasien per hari.

Adapun data pasien yang berobat hanya dijadikan arsip dan untuk melihat riwayat penyakit pasien.

Dalam data mining teknik clustering data penyakit pasien dapat dijadikan informasi baru yang berguna bagi pihak puskesmas atau pihak terkait sebagai bahan penyuluhan kepada masyarakat. Salah satu data yang perlu diperhatikan pengelompokkan penyakit berbasis lingkungan dan bukan berbasis lingkungan.

Untuk mengelompokkan penyakit berbasis lingkungan ataupun tidak dapat menggunakan metode clustering. Fungsi dari pengelompokkan adalah mengurangi fungsi objektif, meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok. (Syahputra et al., 2020)

Dalam penelitian sebelumnya, algoritma k-means dengan teknik clustering sudah sering digunakan salah satu nya untuk kasus pengelompokkan penyakit, diantaranya penelitian yang dilakukan oleh Ninaria Purba, Poningsih dan Heru Satria Tambunan menerapkan Algoritma *K-Means Clustering* Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau yang menghasilkan daerah memiliki kasus ISPA tertinggi adalah kabupaten Dumai dan Pekanbaru. (Purba et al., 2021), Aninda Rachman, Aden, S.Si., M.Pd dan Yulianti Rusdiana, S.Si., M.Sc dalam penelitiannya penerapan algortima k-means clustering untuk menentukan penyakit menular dan tidak menular, dari 30 sampel data yang diujikan maka diperoleh bahwa penyakit TB paru merupakan penyakit yang paling menular di daerah kecamatan kota Tanggerang

(Rachma et al., 2019). Pengelompokkan penyakit berdasarkan tingkat keparahannya oleh Mentari Tri Indah Rahmayani. Ia melakukan penelitian di daerah Bandar Seikijang hasilnya sebanyak 149 orang dengan persentase 53,59%, penyakit ringan 82 orang dan penyakit berat sebanyak 47 orang. ISPA adalah penyakit yang paling banyak diderita masyarakat yang tinggal disekitar puskesmas disusul oleh penyakit DBD dan malaria. (Rahmayani, 2018).

Keterbaruan dari penelitian ini ialah studi kasus pengelompokkan penyakit kronis lansia warga kelurahan pulogebang, dengan mengembangkan proses klasifikasi ini menggunakan model data mining dengan algoritma k-means.

Dari hasil penelitian dan evaluasi penelitian menunjukkan bahwa algoritma k-means dinilai cukup akurat dalam melakukan pengelompokkan penyakit. Dari analisis di atas, penulis ingin melakukan penelitian dengan mengelompokkan penyakit berbasis lingkungan dan yang bukan berbasis lingkungan dengan teknik clustering algoritma k-means. Tujuannya agar hasil penelitian ini dapat menjadi bahan penyuluhan bagi puskesmas agar masyarakat dapat terhindar dari penyakit yang bersumber dari lingkungan.

Data mining adalah point utama dari proses KDD (*Knowledge Discovery in Database*), yaitu eksplor data dari algoritma, membangun model dari pola yang belum diketahui. (Krisna Ferdinan Leo Simanjuntak, Annita Carolina Br Barus, 2021).

Definisi lain dari Data mining adalah suatu aktifitas untuk mencari pola yang menarik dari data dalam jumlah besar, dan data-data tersebut tersimpan dalam *database*. (Wardhani, 2016)

Untuk pengolahan data dalam skala besar teknik data mining dapat digunakan. Hasil pengolahn data dari teknik data mining ini menghasilkan suatu informasi baru yang berguna untuk mengambil suatu keputusan. (Sugianto et al., 2020)

Teknik clustering dengan algoritma k-means merupakan teknik yang sering digunakan untuk menyelesaikan suatu masalah. (Sugianto et al., 2020), (Gustientiedina et al., 2019), (Utomo, 2020). Proses yang dilakukan dalam metode clustering untuk menyelesaikan masalah dengan menggabungkan data. Data yang digabungkan

mempunyai kesamaan sifat kedalam satu kelompok kecil. Dalam penerapannya agar tujuan yang diinginkan dapat tercapai maka setiap tahapan algoritma teknik clustering harus dilakukan. (Sari et al., 2020)

Pada dasarnya penggunaan algoritma ini tergantung pada data yang didapatkan dan hasil yang ingin dicapai pada akhir proses. (Agustina et al., 2012).

2. METODE PENELITIAN

Pada penelitian ini ada beberapa teknik yang penulis gunakan dalam pengumpulan data, yaitu:

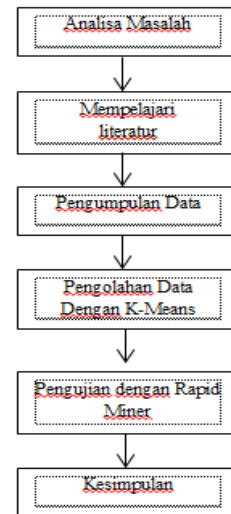
1. Wawancara

Wawancara ini dilakukan kepada salah seorang pegawai puskesmas Sungai Tarab 2 bagian sanitarian, pada bulan Oktober 2021 minggu ketiga. Adapun hasil wawancara nya adalah bahwasanya pasien yang mengidap penyakit berdasarkan faktor lingkungan masih banyak terjadi hal ini dikarenakan masih banyak masyarakat yang belum memahami dampak dari lingkungan yang kurang bersih.

2. Studi kepustakaan

Dalam studi kepustakaan penulis pengumpulan data dengan mempelajari jurnal-jurnal terkait algoritma *k-means*, buku panduan data mining, atikel-artikel serta sumber lainnya yang dianggap perlu dan mendukung Analisa Data.

Di bawah ini merupakan tahapan yang dilakukan dalam penelitian.



Gambar 1. Metode Penelitian

Dari gambar 1 di atas maka dapat penulis jelaskan bahwa:

1. Analisa masalah

Melakukan pengelompokan penyakit berdasarkan lingkungan dan penyakit yang bukan berdasarkan lingkungan dengan menggunakan beberapa variabel yaitu umur dan kode diagnose.

2. Mempelajari Literatur

Adapun literature yang penulis gunakan dalam penelitian ini adalah jurnal-jurnal tentang *k-means*, artikel di internet serta buku referensi yang dapat mendukung dalam penelitian ini.

3. Pengumpulan Data

Pada tahap selanjutnya adalah pengumpulan data. Penulis memperoleh data penelitian ini dari wawancara dengan salah satu pegawai puskesmas bagian sanitarian ditambah dengan studi kepustakaan.

4. Pengolahan Data dengan K-Means

Setelah melakukan pengumpulan data maka tahap selanjutnya adalah mengolah data yang ada dengan metode *k-means*. Pengolahan data dengan metode *k-means* dilakukan secara manual dengan bantuan ms. Excel. Kemudian hasil perhitungan manual akan di bandingkan dengan hasil perhitungan menggunakan aplikasi *rapid miner*.

5. Pengujian dengan RapidMiner

Tahap ini dilakukan setelah selesai melakukan pengolahan data manual dengan metode *k-means*. Tujuan nya adalah untuk

melakukan pengujian dengan aplikasi RapidMiner, apakah sama hasil yang diperoleh perhitungan manual dengan metode k-means dengan aplikasi rapidMiner.

6. Kesimpulan

Setelah pengujian selesai maka akan dilanjutkan dengan proses analisa yang menghasilkan suatu kesimpulan dari penelitian ini.

3. HASIL DAN PEMBAHASAN

3.1 IMPLEMENTASI ALGORITMA K-MEANS

Algoritma k-means digunakan untuk mengelompokkan data pasien kedalam beberapa kelompok, dalam penelitian ini ada dua kategori pengelompokkan penyakit. Pengelompokkan tersebut berdasarkan atribut Umur dan Kode diagnose. Adapun langkah-langkah dari pengelompokkan data adalah sebagai berikut:

1. Langkah pertama adalah menentukan pusat cluster.

Untuk tahap pertama penentuan pusat cluster dapat dilakukan secara acak tanpa ada ketentuannya.

Data yang digunakan pada penelitian ini adalah data pasien yang berobat di puskesmas sungai tarab 2 per oktober 2021 dan diambil 25 sampel data yang akan diolah menggunakan algoritma k-means. Untuk data pasien dapat dilihat pada tabel 1 di bawah ini:

Tabel 1. Sampel Data Penyakit Pasien Puskesmas Sungai Tarab 2

Data Pasien	Jenis Kelamin	Umur (Tahun)	Kode Diagnosa	Diagnosa
1	Perempuan	65	13	Malaria
2	Perempuan	49	13	Malaria
3	Perempuan	77	18	DBD
4	Perempuan	58	16	Dermatitis
5	Perempuan	64	18	DBD
6	Laki-Laki	26	19	Bronkitis
7	Perempuan	19	17	Vertigo
8	Laki-Laki	42	16	Dermatitis
9	Perempuan	47	16	Dermatitis
10	Laki-Laki	8	19	Bronkitis
11	Perempuan	58	18	DBD
12	Laki-Laki	61	13	Malaria
13	Laki-Laki	69	18	DBD
14	Perempuan	30	19	Bronkitis
15	Laki-Laki	24	17	Vertigo

Dari tabel 1 di atas data pasien 3 dan data pasien 14 dijadikan titik pusat cluster awal yang dipilih secara random, seperti pada tabel 2 di bawah ini:

Tabel 2. Titik Pusat Awal Cluster

Data Pasien	Umur	Kode Diagnosa
3	77	18
14	30	19

2. Tahap kedua menghitung jarak setiap data terhadap setiap pusat cluster.

Setelah menentukan cluster awal langkah selanjutnya menghitung jarak setiap data yang ada terhadap setiap pusat cluster.

Dapat menggunakan rumus di bawah ini:

$$de = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (1)$$

Keterangan:

(x ,y) = Koordinat objek

(s ,t) = Koordinat Centroid

i = Banyaknya objek

Dari rumus di atas akan dilakukan perhitungan nilai matrik jarak sebagai berikut :

Jarak Data Pasien 1 ke pusat *cluster*

$$C1 = \sqrt{(65-77)^2 + (13-18)^2}$$

$$C1 = 17$$

$$C2 = \sqrt{(65-30)^2 + (13-19)^2}$$

$$C2 = 29$$

Jarak Data Pasien 2 ke pusat *cluster*

$$C1 = \sqrt{(49-77)^2 + (13-18)^2}$$

$$C1 = 33$$

$$C2 = \sqrt{(49-30)^2 + (13-19)^2}$$

$$C2 = 13$$

Jarak Data Pasien 3 ke pusat *cluster*

$$C1 = \sqrt{(77-77)^2 + (18-18)^2}$$

$$C1 = 0$$

$$C2 = \sqrt{(77-30)^2 + (18-19)^2}$$

$$C2 = 46$$

Jarak Data Pasien 4 ke pusat *cluster*

$$C1 = \sqrt{(58-77)^2 + (16-18)^2}$$

$$C1 = 21$$

$$C2 = \sqrt{(58-30)^2 + (16-19)^2}$$

$$C2 = 25$$

Jarak Data Pasien 5 ke pusat *cluster*

$$C1 = \sqrt{(64-77)^2 + (18-18)^2}$$

$$C1 = 13$$

$$C2 = \sqrt{(64-30)^2 + (18-19)^2}$$

C2= 33

Lanjutkan menghitung untuk data pasien 6 sampai pada data ke-15 sesuai dengan sampel data yang akan diujikan.

3. Tahap ketiga adalah melakukan pengelompokan data.

Setelah selesai menghitung jarak pada setiap data maka langkah selanjutnya adalah melakukan pengelompokan data. Untuk pengelompokan data langkah awal dengan mencari nilai terkecil dari pusat cluster. Data yang memiliki nilai terkecil berada pada kelompok yang sama dengan pusat clusternya. Misalkan untuk data pasien 1, jarak terkecil diperoleh pada cluster pertama, sehingga data pasien 1 akan menjadi anggota dari cluster pertama. Posisi cluster selengkapnya dapat dilihat pada tabel 3 dibawah ini:

Tabel 3. Pengelompokan Data berdasarkan Cluster Terdekat

Data Pasien	C1	C2
1	1	
2		1
3	1	
4	1	
5	1	
6		1
7		1
8		1
9		1
10		1
11	1	
12	1	
13	1	
14		1
15		1

Berdasarkan matrik yang didapatkan pada tabel di atas maka didapatkan pengelompokan sebagai berikut :

C1: (1,3,4,5,11,12 dan 13)

C2: (2,6,7,8,9,10,14 dan 15)

4. Tahap keempat adalah menghitung pusat cluster baru.

Setelah tahap pengelompokan data selesai dikerjakan maka tahap selanjutnya adalah menghitung pusat cluster baru dengan menggunakan rumus

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

Keterangan:

V_{ij} = Centroid rata – rata pada Cluster ke – i untuk variabel ke –j

N_i = Jumlah anggota Cluster ke –i

i,k = Indeks dari Cluster

j = Indeks variabel

X_{kj} = Nilai data ke –k variabel ke – j untuk Cluster tersebut

Dengan menggunakan rumus di atas maka akan dicari pusat cluster baru, seperti perhitungan di bawah ini:

$$C1 = \left(\frac{(65+77+58+64+58+61+69)}{7}; \left(\frac{(13+18+16+18+18+13+18)}{7} \right) \right);$$

$$C1 = (64,6 ; 16,3)$$

$$C2 = \left(\frac{(49+26+19+42+47+8+30+24)}{8}; \left(\frac{(13+19+17+16+16+19+19+17)}{8} \right) \right);$$

$$C2 = (30,6 ; 17,0)$$

Dari perhitungan di atas maka didapatkan pusat cluster baru dalam matrik tabel sebagai berikut :

Tabel 4. Titik Pusat Cluster Baru

C1	64,6	16,3
C2	30,6	17,0

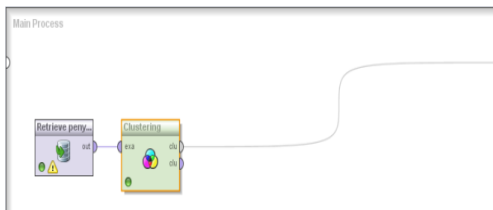
5. Tahap kelima adalah mengulang tahap kedua yaitu menghitung jarak setiap data terhadap setiap cluster.

Setelah memperoleh pusat cluster baru maka lakukan kembali perhitungan mencari jarak setiap data. Langkah ini dilakukan sampai tidak ada lagi anggota cluster yang berubah. Jika masih terjadi perubahan posisi cluster maka ulang lagi mencari pusat cluster baru dan menghitung jarak pada setiap data dan jika posisi cluster sudah tidak mengalami perubahan maka perhitungan dapat dihentikan.

3.2 IMPLEMENTASI RAPID MINER

Dalam implementasinya rapid miner digunakan untuk membandingkan hasil perhitungan manual dengan algoritma k-means dan perhitungan menggunakan aplikasi rapid

miner. Berikut adalah pengolahan data penyakit dengan metode k-means menggunakan aplikasi rapid miner



Gambar 2. Pemodelan k-means dengan rapid miner

Pada gambar 2 di atas merupakan pemodelan k-means clustering menggunakan rapid miner, yang memiliki jumlah cluster sebanyak 2 buah.

Cluster Model

```
Cluster 0: 10 items
Cluster 1: 5 items
Total number of items: 15
```

Gambar 3. Hasil data cluster K-Means dalam implementasi Rapid Miner

Pada gambar 3 merupakan cluster model, sesuai dengan inisialisasi cluster sebanyak 2 buah maka pendefinisannya k dengan jumlah cluster 0 ada 10 items dan cluster 1 ada 5 items dengan total items sebanyak 15 sesuai dengan jumlah sampel data yang digunakan.

☐ Text View	☐ Folder View	☐ Graph View	☒ Centroid Table
Attribute	cluster_0	cluster_1	
Umur (Tahun)	59	21.400	
Kode Diagnosa	15.900	18.200	

Gambar 4. Data centroid berdasarkan cluster dan atribut

Pada gambar 4 di atas merupakan perhitungan antara jarak cluster dan centroid dengan rapid miner. Dapat dilihat pada cluster 0 untuk attribute umur memiliki nilai 59 dan 21,4 untuk nilai cluster_1 sedangkan pada attribute kode diagnose cluster 0 memiliki nilai 15,9 dan cluster 1 nilainya 18,2

4. SIMPULAN

Dari hasil penelitian yang dilakukan maka dapat diambil beberapa kesimpulan, yaitu:

1. Penelitian ini bertujuan untuk mengelompokkan penyakit berdasarkan lingkungan dan yang bukan berdasarkan lingkungan dengan menggunakan variabel umur dan kode diagnose.
2. Inisialisasi cluster sebanyak 2 buah, yaitu cluster 0 dan cluster 1. Dimana cluster 0 mengelompokkan penyakit berbasis lingkungan sedangkan cluster 1 tidak berbasis lingkungan. Dimana penyakit yang termasuk cluster_0 adalah malaria, dbd dan dermatitis sebanyak 10 items sedangkan penyakit yang termasuk cluster_1 adalah vertigo dan bronkitis sebanyak 5 items.
3. Dari perhitungan manual teknik k-means memperoleh hasil yang sama dengan menggunakan aplikasi rapid miner, pembuktian ini mencapai nilai keakuratan 90% dari sampel data yang dilakukan.

5. DAFTAR PUSTAKA

- Agustina, S., Yhudo, D., Santoso, H., Marnasusanto, N., Tirtana, A., & Khusus, F. (2012). Clustering Kualitas Beras Berdasarkan Ciri Fisik Menggunakan Metode K-Means Algoritma. *Clustering K-Means*.
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Irmawati, S. (2017). Kualitas Pelayanan Kesehatan Di Puskesmas Sangurara Kecamatan Tatanga Kota Palu. *Katalogis*, 5(1), 188–197.
- Krisna Ferdinan Leo Simanjuntak, Annita Carolina Br Barus, A. (2021). Implementasi Metode Decision Tree Dan Algoritma C4.5 Untuk Klasifikasi Kepribadian Masyarakat. *JOISIE Journal Of Information System And Informatics Engineering*, 5(1), 51–59.

- Purba, N., Poningsih, P., & Tambunan, H. S. (2021). Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau. *Journal of Information System Research (JOSH)*, 2(3), 220–226. <http://ejurnal.seminar-id.com/index.php/josh/article/view/736>
- Rachma, A., Aden, A., & Rusdiana, Y. (2019). Analisis Cluster Menggunakan Algoritma K-Means Cluster Untuk Culstering Jenis Penyakit Menular Pada Puskesmas Di Kecamatan Kota Tangerang. *Jurnal Saintika Unpam: Jurnal Sains Dan Matematika Unpam*, 2(1), 15. <https://doi.org/10.32493/jsmu.v2i1.2915>
- Rahmayani, M. T. I. (2018). Analisis Clustering Tingkat Keparahan Penyakit Pasien Menggunakan Algoritma K-Means. *Jurnal Inovasi Teknik Informatika*, 1(2), 40–44.
- Sari, Y. P., Primajaya, A., & Irawan, A. S. Y. (2020). Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang. *INOVTEK Polbeng - Seri Informatika*, 5(2), 229. <https://doi.org/10.35314/isi.v5i2.1457>
- Sugianto, C. A., Rahayu, A. H., & Gusman, A. (2020). Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Cigugur Tengah. *Journal of Information Technology*, 2(2), 39–44. <https://doi.org/10.47292/joint.v2i2.30>
- Syahputra, S., Ramadani, S., Manaor, A., & Pardede, H. (2020). Menentukan Strategi Promosi Menggunakan Algoritma Clustering K-Means. *JOISIE Journal Of Information System And Informatics Engineering*, 4(1), 7–14.
- Utomo, W. (2020). Penerapan Algoritma K-Means Untuk Pengelompokan Penyakit Kronis pada Warga Lansia (Studi Kasus Pada: Posyandu Lansia RW 07). *Jurnal Media Informatika ...*, 4(4), 1153–1161. <https://doi.org/10.30865/mib.v4i4.2410>
- Wardhani, A. K. (2016). K-Means Algorithm Implementation for Clustering of Patients Disease in Kajen Clinic of Pekalongan. *Jurnal Transformatika*, 14(1), 30. <https://doi.org/10.26623/transformatika.v14i1.387>