

Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Cigugur Tengah

Castaka Agus Sugianto¹, Ayu Hendrati Rahayu², Aditia Gusman³

^{1,3}Program Studi Teknik Informatika - Politeknik TEDC Bandung

²Program Studi Rekam Medik dan Informasi Kesehatan - Politeknik TEDC Bandung
Jl. Politeknik-Pesantren KM2 Cibabat Cimahi Utara – Cimahi - Jawa Barat - Indonesia

¹castaka@poltektdc.ac.id

²ayuhendrati@poltektdc.ac.id

³aditiagusman19@gmail.com

Intisari—Puskesmas Cigugur Tengah dalam setiap harinya melayani pasien sekitar 150 orang dari berbagai wilayah daerah Cigugur Tengah. Dengan bertambahnya jumlah pasien tersebut, maka bertambah pula data pasien setiap harinya, sehingga sejumlah data tidak dapat dipelajari lebih lanjut dan data tersebut hanya digunakan sebagai arsip saja. Berdasarkan latar belakang tersebut, maka penulis ingin mengolah data tersebut untuk mengelompokan penyakit pasien berdasarkan penyakit akut dan penyakit tidak akut menggunakan teknik data mining dengan metode *clustering* dengan algoritma *k-means* dan algoritma *k-medoids* sebagai banding. Sehingga nantinya dapat membantu pihak Puskesmas Cigugur Tengah untuk mengetahui penyakit apa yang paling banyak diderita pasien, kemudian dapat membantu pihak pemerintah khususnya Dinas Kesehatan dalam pemberian penyuluhan kesehatan kepada masyarakat sekitar. Berdasarkan hasil pengujian dari algoritma *k-means* dan algoritma *k-medoids*, didapat *cluster model* untuk algoritma *k-means* sebanyak 241 items pada *cluster_0* atau penyakit akut sebesar 96% dan 9 items pada *cluster_1* atau penyakit tidak akut sebesar 4% dari 250 data, sedangkan untuk algoritma *k-medoids* sebanyak 224 items pada *cluster_0* atau penyakit akut sebesar 90% dan 26 items pada *cluster_1* atau penyakit tidak akut sebesar 10% dari 250 data, maka penyakit yang paling banyak diderita pasien pada Puskesmas Cigugur Tengah adalah penyakit akut sebesar 93%, dengan nilai *Davies Bouldin* untuk algoritma *k-means* sebesar -0.453 dan algoritma *k-medoids* sebesar -1.276. Dari hasil tersebut dapat dikatakan bahwa algoritma yang menghasilkan nilai *Davies Bouldin* terkecil dianggap sebagai algoritma yang lebih baik, maka dapat disimpulkan bahwa algoritma *k-means* lebih baik dari algoritma *k-medoids* yang menghasilkan nilai rata-rata *Davies Bouldin* sebesar -1.276.

Kata kunci— Data Mining, Clustering, Algoritma K-Means, Algoritma K-Medoids, Penyakit Akut dan Penyakit Tidak Akut.

Abstract— The Central Cigugur Community Health Center serves around 150 patients every day from various regions in the Central Cigugur area. Along with the increase in the number of patients, the patient's data is increased every day. So that a number of data cannot be studied further and the data is only used as an archive. Based on this background, the authors want to process the data to classify the patient's disease based on acute disease and non-acute disease using data mining techniques with clustering method with *k-means* algorithm and *k-medoids* algorithm as a comparison. So that later it can help the Cigugur Tengah Health Center to find out what diseases most suffer from patients, then can help the government especially the Health Service in providing health counseling to the surrounding community. Based on the test results of the *k-means* algorithm and the *k-medoids* algorithm, the cluster model for 241 items of *cluster-0* or 96% of the *k-means* algorithm was obtained and 9 items for *cluster_1* or non-acute illness for 4% of 250 data , whereas for the *k-medoids* algorithm as much as 224 items in *cluster_0* or acute disease by 90% and 26 items in *cluster_1* or non-acute diseases in the amount of 10% of 250 data, the most common disease suffered by patients at Cigugur Tengah Health Center is acute disease by 93%, with the *Davies Bouldin* value for the *k-means* algorithm of -0.453 and the *k-medoids* algorithm of -1.276. From these results it can be said that the algorithm that produces the smallest *Davies Bouldin* value is considered a better algorithm, so it can be concluded that the *k-means* algorithm is better than the *k-medoids* algorithm which produces the average *Davies Bouldin* value of -1.276.

Keywords— Data Mining, Clustering, K-Means Algorithm, K-Medoids Algorithm, Acute Disease and Non-Acute Disease.

I. PENDAHULUAN

Puskesmas merupakan salah satu instansi kesehatan yang berada di tingkat kecamatan, termasuk Puskesmas Cigugur Tengah yang berada di kecamatan Cimahi Tengah. Dalam rangka menciptakan pelayanan Instansi Kesehatan yang baik diperlukan tata kerja yang tertib, rapi, dan teliti sehingga menghasilkan informasi yang cepat, akurat, dan tepat waktu sesuai kebutuhan. Seperti instansi kesehatan pada umumnya, Puskesmas Cigugur Tengah juga memberikan pelayanan kesehatan bagi masyarakat setiap harinya. Setiap harinya, Puskesmas Cigugur Tengah melayani pasien sekitar 150 orang dari berbagai wilayah yang ada di daerah Cigugur Tengah.

Seiring dengan bertambahnya jumlah pasien tersebut, maka bertambah pula data pasien pada setiap harinya,

sehingga dari data yang banyak tersebut tidak dapat dipelajari lebih lanjut dan data tersebut hanya digunakan sebagai arsip saja. Salah satu data yang perlu diperhatikan yaitu mengenai penyakit akut dan tidak akut yang banyak diderita pasien pada Puskesmas Cigugur Tengah. Sehingga nantinya dapat membantu pihak pemerintah khususnya Dinas Kesehatan untuk mengetahui penyakit apa yang paling banyak diderita dan juga bisa menjadi bahan acuan dalam pemberian penyuluhan kesehatan kepada masyarakat sekitar.

Untuk mempermudah proses pengelolaan data yang banyak, suatu Instansi di perlukan suatu sistem dalam mengambil keputusan untuk mengetahui pengelompokan penyakit berdasarkan penyakit akut dan tidak akut yang banyak diderita pasien pada Puskesmas Cigugur Tengah. Salah satunya yaitu menggunakan teknik *Data Mining*.

Sehingga data yang banyak tersebut dapat dimanfaatkan secara optimal, kemudian hasilnya dapat dijadikan bahan atau dasar penuluhan kesehatan oleh Dinas Kesehatan Kota Cimahi.

Data mining merupakan metode yang digunakan dalam pengolahan data berskala besar. Data yang diolah dengan teknik data mining ini kemudian menghasilkan suatu pengetahuan baru, hasil dari pengolahan data tersebut dapat digunakan dalam menentukan keputusan. Dalam data mining juga terdapat metode - metode yang dapat digunakan seperti *Classification*, *Clustering*, *Estimation*, *Prediction*, *Association* [1].

Clustering merupakan metode yang digunakan dalam data mining yang cara kerjanya mencari dan mengelompokkan data yang mempunyai kemiripan karakteristik antara data satu dengan data lainnya yang telah diperoleh [1]. Teknik *Clustering* yang paling sering digunakan dalam data mining yaitu algoritma *K-Means* dan *K-Medoids* [2][3][4][5][6]. Kelebihan dari algoritma *K-Means* ini mampu meminimalkan jarak antara data ke *clusternya*. Pada dasarnya penggunaan algoritma ini dalam proses *clustering* tergantung pada data yang didapatkan dan yang ingin dicapai diakhir proses [7].

Beberapa penelitian terutama *clustering* sering digunakan untuk pengelompokan penyakit diantaranya penelitian yang di lakukan oleh Ade Bastian, Harun Sujadi, dan Gigin Febrianto menerapakan algoritma k-means clustering analysis pada penyakit menular manusia cluster terbanyak yaitu penyakit diare[8], yang di lakukan Anindya Khrisna wardhani Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan hasilnya 376 penyakit akut dan 424 penyakit tidak akut dari total data 1000 [9], Parasian D P Silitonga dan Irene Sri dalam penelitiannya Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering dari sejumlah pasien yang ada, persentasi usia pasien paling tinggi adalah pasien dengan usia tua dan kemudian pasien dengan usia parobaya[10]. Sedangkan penelitian yang dilakukan Mentari Tri Indah Rahmayani terkait Analisis Clustering Tingkat Keparahan Penyakit Pasien Menggunakan Algoritma K-Means hasilnya masyarakat disekitar puskesmas Bandar Seikjang lebih banyak menderita penyakit sedang (C2) dengan jenis penyakit dominan Infeksi Saluran Pernafasan Akut (ISPA), demam berdarah dan malaria[11], penelitian lain oleh Fina Nasari dan Charles Jhony Manto Sianturi terkait Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat Hasilnya *cluster* 1 yaitu Kecamatan Batang Serangan, Brandan Barat dan Permata Jaya sebagai dan Kecamatan Hinai dan Sei Bingai menjadi pusat *cluster* 2 [12].

Berdasarkan latar belakang di atas, maka penulis ingin mengelompokkan penyakit pasien kedalam kelompok penyakit akut dan penyakit tidak akut menggunakan teknik data mining dengan metode *clustering* dengan algoritma *k-means* dan algoritma *k-medoids* sebagai banding.

II. METODE PENELITIAN

Dalam penelitian ini Ada 8 (delapan) tahapan yang akan dilakukan diantaranya:

A. Menentukan Masalah

Penentuan masalah yang penulis lakukan merupakan proses awal dalam penelitian ini. Proses ini sudah berlangsung di jelaskan pada latar belakang mengenai pengelompokan penyakit pasien kedalam penyakit akut dan tidak akut menggunakan algoritma *K-Means* algoritma *K-Medoids* sebagai banding.

B. Menentukan Tujuan dan Ruang Lingkup

Menentukan tujuan dan ruang lingkup merupakan apa yang ingin di capai oleh penulis dalam penelitian serta batasan yang di pakai agar penelitian tidak terlalu melebar.

C. Mencari Literature

Mencari literature atau referensi dilakukan untuk mencari berbagai sumber yang berkaitan tentang data mining, clustering, dan algoritma *k-means*. Adapun metode yang digunakan yaitu melakukan Studi Pustaka

Dalam penelitian ini menggunakan metode studi pustaka yang merupakan sebuah metode yang dilakukan dengan cara mengumpulkan dan mencari sumber dari buku, internet, jurnal yang berkaitan dengan penelitian yang dilakukan. Data - data yang dikumpulkan oleh penulis dari metode studi pustaka adalah :

1. Materi tentang data mining.
2. Materi tentang penyakit akut dan tidak akut.
3. Materi tentang rapidminer.
4. Mengumpulkan referensi jurnal yang memiliki korelasi dengan data mining, clustering, algoritma *k-means* dan algoritma *k-medoids*.
5. Mengambil dataset penyakit pasien.
6. Teori – teori yang digunakan pada penelitian ini.

D. Pengumpulan Data

Proses pengumpulan data mempunyai peranan yang penting dalam penelitian ini, karena metode pengumpulan data akan menentukan kualitas dan akurasi data yang akan dikumpulkan selama proses penelitian dengan berbagai macam metode pengumpulan data, penulis akan menggunakan metode pengumpulan data sebagai berikut :

1. Data Primer

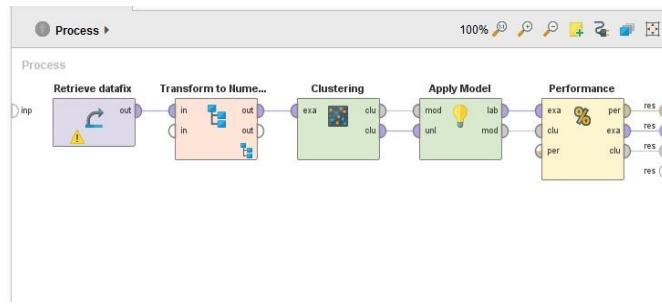
Data primer adalah data yang diperoleh secara langsung dari sumber data tersebut yang berhubungan dengan penelitian yang dilakukan. Pada penelitian ini data yang digunakan diperoleh langsung dari Puskesmas Cigugur Tengah.

2. Data Sekunder

Data sekunder adalah data yang diperoleh secara tidak langsung, biasanya dari pihak kedua yang mengolah data untuk orang lain. Data sekunder dalam penelitian ini diperoleh dari *jurnal*, *paper*, artikel yang berkaitan dengan Data Mining, Clustering, Algoritma K-Means dan Algoritma K-Medoids.

3. Observasi

Observasi merupakan metode pengumpulan data dengan cara mengamati secara langsung kejadian dilapangan. Dalam penelitian ini melakukan observasi di Puskesmas Cigugur Tengah yang



Gambar 4. Pemodelan Clustering K-Means

B. Hasil Data Cluster K-Means

Dengan menggunakan pemodelan *k-means clustering* seperti gambar 4 diatas, dengan inisialisasi jumlah *cluster* sebanyak 2 buah, maka didapatkan hasil dengan *cluster* yang terbentuk adalah 2, sesuai dengan pendefinisiannya dengan jumlah *cluster* 0 ada 241 item, *cluster* 1 ada 9 item dengan total jumlah data sebanyak 250 adapun penjelasanya bisa dilihat pada gambar 5.

Cluster Model

```

Cluster 0: 241 items
Cluster 1: 9 items
Total number of items: 250
  
```

Gambar 5. Hasil Data Cluster K-Means

Dimana akurasi berdasarkan rata-rata *within centroid distance* dapat dilihat pada gambar 6.

PerformanceVector

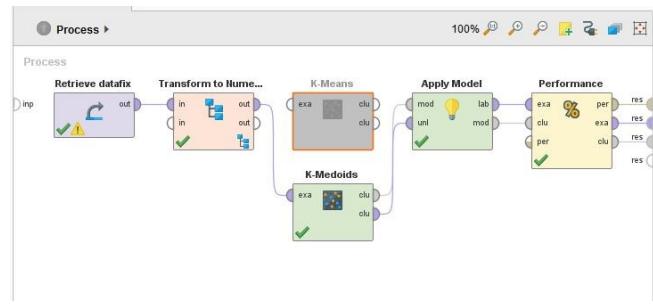
```

PerformanceVector:
Avg. within centroid distance: -43954.608
Avg. within centroid distance_cluster_0: -21401.536
Avg. within centroid distance_cluster_1: -647875.778
Davies Bouldin: -0.453
  
```

Gambar 6. Rata-rata *centroid distance*

C. Pengujian Algoritma K-Medoids Pada RapidMiner

Proses pengujian data ini dilakukan sebagai pembanding tanpa melakukan perubahan apapun pada algoritma yang terdapat pada *rapidminer*. Pengujian dilakukan menggunakan algoritma *k-medoids* seperti digambarkan pada gambar 7.



Gambar 7. Pemodelan Clustering K-Medoids

D. Hasil Data Cluster K-Medoids

Dengan menggunakan pemodelan *k-medoids clustering* seperti gambar 7 diatas, dengan inisialisasi jumlah *cluster* sebanyak 2 buah, maka didapatkan hasil dengan *cluster* yang terbentuk adalah 2, sesuai dengan pendefinisiannya dengan jumlah *cluster* 0 ada 224 item, *cluster* 1 ada 26 item dengan total jumlah data sebanyak 250 adapun penjelasanya bisa dilihat pada gambar 8.

Cluster Model

```

Cluster 0: 224 items
Cluster 1: 26 items
Total number of items: 250
  
```

Gambar 8. Hasil Data Cluster K-Medoids

Dimana akurasi berdasarkan rata-rata *within centroid distance* dapat dilihat pada gambar 9.

PerformanceVector

```

PerformanceVector:
Avg. within centroid distance: -99830.496
Avg. within centroid distance_cluster_0: -4449.094
Avg. within centroid distance_cluster_1: -921577.962
Davies Bouldin: -1.276
  
```

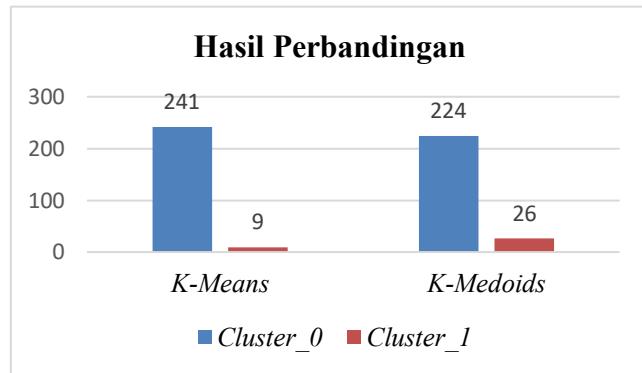
Gambar 9. Rata-rata *centroid distance*

E. Perbandingan Algoritma K-Means dan Algoritma K-Medoids

Algoritma *K-Means* dan Algoritma *K-Medoids* sama-sama menghasilkan dua *cluster* sesuai dengan yang dibutuhkan, dimana *cluster* pertama (C0) merupakan penyakit akut dan *cluster* kedua (C1) merupakan penyakit tidak akut. Hasil perbandingan dari Algoritma *K-Means* dan Algoritma *K-Medoids* dapat dilihat pada tabel 1 dan gambar 10.

Tabel 1 Tabel Hasil Perbandingan *Algoritma K-Means* dan *Algoritma K-Medoids*

Algoritma	Cluster_0	Cluster_1
Algoritma <i>K-Means</i>	241 items	9 items
Algoritma <i>K-Medoids</i>	224 items	26 items



Gambar 10. Grafik Hasil Perbandingan *Algoritma K-Means* dan *Algoritma K-Medoids*

IV. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut :

1. Dalam penelitian ini penulis ingin mengelompokan penyakit pasien kedalam penyakit akut dan tidak akut, maka jumlah *cluster* yang digunakan dalam penelitian ini sebanyak 2 *cluster*, yaitu *cluster_0* dan *cluster_1*. Dimana *cluster_0* untuk mengelompokan penyakit akut, sedangkan *cluster_1* untuk mengelompokan penyakit tidak akut.
2. Berdasarkan pengujian yang telah dilakukan pada *rapidminer* menggunakan *algoritma k-means*, maka didapatkan nilai performa berdasarkan rata – rata *avg. within centroid distance* sebesar -43954.608, *avg. within centroid distance_cluster_0* sebesar -21401.536, *avg. within centroid distance_clsuter_1* sebesar -647875.778 dan *Davies Bouldin* sebesar -0.453.
3. Berdasarkan hasil pengujian dari *algoritma k-means* dan *algoritma k-medoids*, didapat *cluster model* untuk *algoritma k-means* sebanyak 241 items pada *cluster_0* atau penyakit akut dan 9 items pada *cluster_1* atau penyakit tidak akut, sedangkan untuk *algoritma k-medoids* sebanyak 224 items pada *cluster_0* atau penyakit akut dan 26 items pada *cluster_1* atau penyakit tidak akut, dengan nilai *Davies Bouldin* untuk *algoritma k-means* sebesar -0.453 dan *algoritma k-medoids* sebesar -1.276. Dari hasil tersebut dapat dikatakan bahwa *algoritma* yang menghasilkan nilai *Davies Bouldin* terkecil dianggap sebagai *algoritma* yang lebih baik, maka dapat disimpulkan bahwa *algoritma k-*

means lebih baik dari *algoritma k-medoids* yang menghasilkan nilai rata – rata *Davies Bouldin* sebesar -1.276.

REFENSI

- [1] J. O. Ong, "Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university," *J. Ilm. Tek. Ind.*, vol. vol.12, no. no. juni, pp. 10–20, 2013.
- [2] N. L. Anggreini and S. Tresnawati, "Komparasi Algoritma K-Means dan K-Medoids untuk menangani Strategi Promosi di Politeknik TEDC Bandung," *J. TEDC*, vol. 14, no. 2, pp. 120–127, 2020.
- [3] C. A. Sugianto, "Analisis Komparasi Algoritma Klasifikasi Untuk Menangani Data Tidak Seimbang Pada Data Kebakaran Hutan," *Techno.Com*, vol. 14, no. 4, pp. 336–342, 2015, doi: 10.33633/tc.v14i4.992.
- [4] N. L. Anggreini, "Teknik Clustering dengan Algoritma K-Medoids untuk menangani Strategi Promosi di Politeknik TEDC Bandung," *Jurnal Teknol. Inf. dan Pendidik.*, vol. 12, no. 2, pp. 1–7, 2019.
- [5] C. A. Sugianto and T. H. Apandi, "Algoritma K-Means Untuk Menangani Data Tidak Seimbang Pada Data Kebakaran Hutan," in *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan (SEMANTIK)*, 2015, pp. 77–81.
- [6] N. L. Anggreini and I. Budi, "Teknik Clustering Dengan Algoritma K-Means Untuk Menangani Strategi Promosi di Politeknik TEDC Bandung," *J. TEDC*, vol. 13, no. 1, pp. 40–45, 2019.
- [7] S. Agustina, D. Yhudo, H. Santoso, N. Marnasusanto, A. Tirtana, and F. Khusnu, "CLUSTERING KUALITAS BERAS BERDASARKAN CIRI FISIK MENGGUNAKAN METODE K-MEANS Algoritma," *Clust. K-Means*, pp. 1–7, 2012.
- [8] A. Bastian, H. Sujadi, and G. Febrianto, "Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)," *J. Sist. Inf.*, vol. 14, no. 1, pp. 26–32, 2018.
- [9] A. K. Wardhani, "Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan," *J. Transform.*, vol. 14, no. 1, pp. 30–37, 2016.
- [10] P. D. P. Silitonga and I. Sri, "Klusterisasi Pola Penyebaran Penyakit Pasien Berdasarkan Usia Pasien Dengan Menggunakan K-Means Clustering," *TIRES (Technology Informatics Comput. Syst.)*, vol. 6, no. 2, pp. 22–25, 2017.
- [11] M. T. I. Rahmayani, "Analisis Clustering Tingkat Keparahan Penyakit Pasien Menggunakan Algoritma K-Means (Studi Kasus di Puskesmas Bandar Seikijang)," *J. Inov. Tek. Inform.*, vol. 1, no. 2, pp. 40–44, 2018.
- [12] F. Nasari and C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," *CogITO Smart J.*, vol. 2, no. 2, pp. 108–119, 2016, doi: 10.31154/cogito.v2i2.19.108-119.

(halaman ini sengaja dikosongkan)