
PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOID DALAM PENGELOMPOKAN DATA PASIEN BERDASARKAN REKAM MEDIS DI PUSKESMAS M. THAHA BENGKULU SELATAN

Adetri Suprianto^{1,3}, Herlina Latipa Sari², Ricky Zulfiandry³

Universitas Dehasen, Bengkulu

e-mail: ¹adetrissuprianto16@gmail.com, ²herlinalatipasari@unived.ac.id

³ricky.zulfiandry@gmail.com

Abstract: *M. Thaha Public Health Center of South Bengkulu is one of public health centers in South Bengkulu. So far, data management of medical record is still done manually by filling out the medical record form provided by Public Health Center. Because data management is still manual, it is difficult for Public Health Center to choose the right material for outreach to the community about a disease. In addition, the data in Public Health Center is compiled in archive form, so when you need information from the data, it takes quite a long time because you have to sort out the archives that have been stored one by one. The data grouping on the elderly population in K-Means Method and the K-Medoids Method is divided into 2 groups, namely Cluster C1 and Cluster C2. The implementation of grouping patient data based on patient medical records at M. Thaha Public Health Center of South Bengkulu can be used to find out which diseases are included in groups with high or low intensity based on patient medical records and can help Public Health Center in selecting material for outreach to the community, especially about disease. The results of the comparative analysis between K-Means and K-Medoids methods, it was found that the differences in grouping, iteration and processing time results depended on the initial centroid value used in each method. Based on the results of the tests that have been carried out, the patient data grouping application based on patient medical records at M. Thaha Public Health Center of South Bengkulu was successfully carried out, and can provide information based on 2 groups, namely Cluster C1 and Cluster C2, and the functionality of the application has run as expected.*

Keywords: *Comparison, K-Means Algorithm, K-Medoid Algorithm, Patient Data, Medical Records, M. Thaha Public Health Center of South Bengkulu*

Abstrak: Puskesmas M.Thaha Bengkulu Selatan merupakan salah satu pusat kesehatan masyarakat yang terdapat di Bengkulu Selatan. Selama ini pengelolaan data rekam medis masih dilakukan secara manual dengan mengisi form rekam medis yang telah disediakan dari Puskesmas. Dikarenakan pengelolaan data yang masih manual, membuat pihak puskesmas kesulitan dalam memilih materi yang tepat untuk penyuluhan suatu penyakit ke masyarakat. Selain itu data-data yang ada di Puskesmas tersebut disusun dalam bentuk arsip, sehingga ketika membutuhkan suatu informasi dari data tersebut, dibutuhkan waktu yang cukup lama dikarenakan harus memilah satu persatu arsip yang telah disimpan. Pengelompokan data penduduk lanjut usia pada Metode K-Means dan Metode K-Medoids dibagi menjadi 2 kelompok yaitu Cluster C1 dan Cluster C2. Aplikasi pengelompokan data pasien berdasarkan rekam medis pasien di Puskesmas M. Thaha Bengkulu Selatan dapat digunakan untuk mengetahui penyakit mana yang masuk ke dalam kelompok dengan intensitas tinggi atau rendah berdasarkan rekam medis pasien dan dapat membantu pihak puskesmas dalam memilih materi untuk penyuluhan ke masyarakat khususnya tentang penyakit. Hasil analisis perbandingan antara Metode K-Means dan K-Medoids, diperoleh bahwa perbedaan hasil pengelompokan, iterasi dan waktu proses terjadi tergantung nilai centroid awal yang digunakan pada masing-masing metode. Berdasarkan hasil pengujian yang telah dilakukan, aplikasi pengelompokan data

pasien berdasarkan rekam medis pasien di Puskesmas M. Thaha Bengkulu Selatan berhasil dilakukan, dan dapat memberikan informasi berdasarkan 2 kelompok yaitu Cluster C1 dan Cluster C2, serta fungsional dari aplikasi telah berjalan sesuai dengan yang diharapkan.

Kata kunci: Perbandingan, Algoritma K-Means, Algoritma K-Medoid, Data Pasien, Rekam Medis, Puskesmas M. Thaha Bengkulu Selatan

PENDAHULUAN

Puskesmas merupakan fasilitas pelayanan kesehatan yang menyelenggarakan upaya kesehatan masyarakat dan upaya kesehatan perseorangan tingkat pertama, dengan lebih mengutamakan upaya promotif dan preventif, untuk mencapai derajat kesehatan masyarakat yang setinggi-tingginya. Program kerja puskesmas yaitu meningkatkan layanan kesehatan untuk masyarakat. Puskesmas memiliki program kerja yaitu promosi kesehatan yang kegiatannya berupa penyuluhan kesehatan ke masyarakat. Menerapkan promosi kesehatan dengan topik penyuluhan mengenai suatu penyakit yang tepat bagi masyarakat desa dapat menjadi salah satu upaya terbaik untuk meningkatkan kepercayaan masyarakat terhadap Puskesmas.

Puskesmas M.Thaha Bengkulu Selatan merupakan salah satu pusat kesehatan masyarakat yang terdapat di Bengkulu Selatan. Selama ini pengelolaan data rekam medis masih dilakukan secara manual dengan mengisi form rekam medis yang telah disediakan dari Puskesmas. Dikarenakan pengelolaan data yang masih manual, membuat pihak puskesmas kesulitan dalam memilih materi yang tepat untuk penyuluhan suatu penyakit ke masyarakat. Selain itu data-data yang ada di Puskesmas tersebut disusun dalam bentuk arsip, sehingga ketika membutuhkan suatu informasi dari data tersebut, dibutuhkan waktu yang cukup lama dikarenakan harus memilah satu persatu arsip yang telah disimpan.

Oleh karena itu dalam penelitian dilakukan pengembangan sistem dengan membuat aplikasi yang digunakan untuk

mengelompokkan penyakit berdasarkan hasil pengolahan rekam medis pasien. Hal ini dapat membantu pihak puskesmas dalam memilih materi untuk penyuluhan ke masyarakat dikarenakan hasil pengelompokan dibagi menjadi 2 kelompok yaitu kelompok tinggi dan kelompok rendah dengan melihat jumlah pasien yang terindikasi penyakit. Untuk mempermudah proses pengelompokan data tersebut, maka pada aplikasi telah diterapkan 2 algoritma clustering yaitu K-Means dan K-Medoid. Selain itu, pada penelitian ini dilakukan perbandingan terhadap 2 algoritma clustering tersebut berdasarkan 3 aspek yaitu hasil pengelompokan, jumlah iterasi, dan waktu proses.

METODE

Penelitian ini menggunakan perbandingan dua metode yaitu K-Means dan K-Medoid. K-means merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok (Maulida, 2018:170). Menurut Kusuma (2020 :145), Medoid adalah anggota

cluster yang menjadi titik pusat cluster. Jumlah medoid di dalam populasi adalah sama dengan k. Proses K-Medoid Clustering terdiri dari dua tahap. Tahap pertama adalah tahap inisialisasi yaitu penentuan posisi-posisi medoid untuk pertama kali. Tahap kedua adalah tahap iterasi yaitu reposisi medoid dan perhitungan jarak total. Tahap kedua berakhir saat jarak total atau nilai y mencapai nilai minimum.

Data Rekam Medis Pasien

Adapun data rekam medis pasien diperoleh dari Puskesmas M. Thaha Bengkulu Selatan. Setelah dikelompokkan maka data seperti pada table berikut ini.

Tabel 1. Olah Data Penyakit Berdasarkan Rekam Medis

Penyakit	Jumlah Pasien
Abdominal discofat	2
<i>Abdominal pain</i>	1
Anorexia	1
CC	1
Cefalgia	1
Costokondritis	1
Dislipenden	3
DM	1
DM Terkontrol	2
Dyspepsia	6
F holder	1
G2 P1 A0	1
Hemoroid	4
HT	11
HT gr TI	1
HT terkontrol	1
LBP	3
Neuralgia	1
Nyeri badan	1
Obs febris	1
Os Febris	1
Otitis Eksternal	2
Papilloma	2
VL	1

Penerapan Metode K-Means

Adapun langkah-langkah pada Metode K-Means di antaranya :

1. Pilih jumlah klaster
2. Inisialisasi awal dan pusat klaster dilakukan secara *random*, Adapun nilai centroid awal pada masing-masing cluster
3. Setiap data ditempatkan ke pusat klaster terdekat berdasarkan jarak antar objek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan metode jarak *euclidean* dengan rumus :

$$d_{(x,y)} = \sqrt{\sum_{j=1}^n (x_i - y_i)^2}$$

4. Hitung pusat klaster yang baru dengan keanggotaan yang baru dengan cara menghitung rata-rata objek pada klaster
5. Setiap data ditempatkan ke pusat klaster terdekat berdasarkan jarak antar objek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan metode jarak *euclidean* dengan rumus :

$$d_{(x,y)} = \sqrt{\sum_{j=1}^n (x_i - y_i)^2}$$

6. Hitung pusat klaster yang baru dengan keanggotaan yang baru dengan cara menghitung rata-rata objek pada klaster. penghitungan bisa juga dengan menggunakan median.
7. Setiap data ditempatkan ke pusat klaster terdekat berdasarkan jarak antar objek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan metode jarak *euclidean* dengan rumus :

$$d_{(x,y)} = \sqrt{\sum_{j=1}^n (x_i - y_i)^2}$$

8. Hitung pusat kluster yang baru dengan keanggotaan yang baru dengan cara menghitung rata-rata objek pada kluster. penghitungan bisa juga dengan menggunakan median
9. Setiap data ditempatkan ke pusat kluster terdekat berdasarkan jarak antar objek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan metode jarak *euclidean* dengan rumus :

$$d_{(x,y)} = \sqrt{\sum_{j=1}^n (x_i - y_i)^2}$$

10. Hitung pusat kluster yang baru dengan keanggotaan yang baru dengan cara menghitung rata-rata objek pada kluster. penghitungan bisa juga dengan menggunakan median
11. Setiap data ditempatkan ke pusat kluster terdekat berdasarkan jarak antar objek. Pada tahap ini jarak dihitung dengan menentukan kemiripan atau ketidakmiripan data dengan metode jarak *euclidean* dengan rumus :

$$d_{(x,y)} = \sqrt{\sum_{j=1}^n (x_i - y_i)^2}$$

12. Proses iterasi berhenti dikarenakan tidak terjadi perubahan data pada masing-masing cluster, sehingga menghasilkan nilai akhir pada iterasi keempat

Penerapan Metode K-Medoids

Adapun langkah-langkah Metode *K-Medoid clustering* di antaranya :

1. Inisialisasi pusat cluster sebanyak k (jumlah cluster) Jumlah kluster yang digunakan sebanyak 2 Cluster
2. Alokasikan setiap data (objek) ke kluster terdekat menggunakan persamaan ukuran jarak *Euclidean Distance* dengan persamaan :

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)}$$

3. Pilih secara acak objek pada masing-masing kluster sebagai kandidat *medoid* baru
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoid* baru

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)}$$

5. Hitung total simpangan (S). Setelah didapatkan nilai jarak iterasi ke-1 dan iterasi ke-2, hitung total simpangan (S) dengan mencari selisih dari total cost baru dengan nilai total cost lama. Dengan ketentuan jika $S < 0$, maka tukar nilai objek dengan menentukan medoid baru.

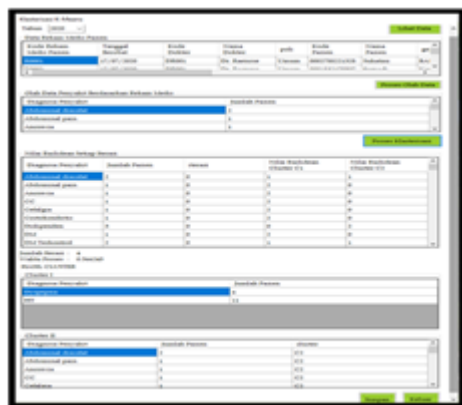
Aspek Perbandingan Metode K-Means dan Metode K-Medoid Aspek yang dibandingkan terdiri dari 3 yaitu hasil pengelompokan, jumlah iterasi yang terjadi, dan waktu proses pengelompokan

HASIL DAN PEMBAHASAN

Aplikasi pengelompokan data pasien berdasarkan rekam medis pasien di Puskesmas M. Thaha Bengkulu Selatan dibuat menggunakan bahasa pemrograman *Visual Basic .Net* dan *Database SQL Server 2008* dengan menerapkan 2 Metode data mining adalah *K-Means* dan *K-Medoid*. Pengelompokan dilakukan berdasarkan data rekam medis pasien yang diperoleh dari Puskesmas M. Thaha Bengkulu Selatan.

Aplikasi pengelompokan data pasien berdasarkan rekam medis pasien di Puskesmas M. Thaha Bengkulu Selatan dapat digunakan untuk mengetahui penyakit mana yang masuk ke dalam kelompok dengan intensitas tinggi atau rendah berdasarkan rekam medis pasien dan dapat membantu pihak puskesmas dalam memilih materi untuk penyuluhan ke masyarakat khususnya tentang penyakit.

Proses klusterisasi data dilakukan dengan membagi data berdasarkan 2 kelompok yaitu Cluster I dan Cluster II. Selain itu, dilakukan analisis perbandingan terhadap Metode K-Means dan Metode K-Medoid berdasarkan 3 aspek yaitu hasil pengelompokan, jumlah iterasi yang terjadi, dan waktu proses pengelompokan. Setelah melakukan proses klusterisasi menggunakan 2 metode tersebut, diperoleh grafik hasil klusterisasi kedua metode dan juga hasil perbandingan berdasarkan hasil pengelompokan, jumlah iterasi dan waktu proses.

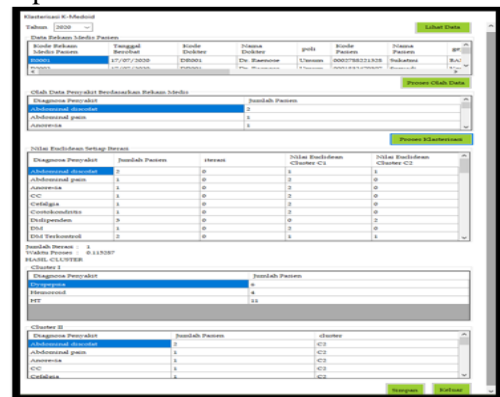


Gambar Klusterisasi Algoritma K-Means

Klusterisasi Algoritma K-Means merupakan antarmuka aplikasi klusterisasi rekam medis pasien yang digunakan untuk menjalankan proses pengelompokan data rekam medis pasien menjadi 2 kelompok yaitu Cluster C1 dan Cluster C2. Dalam proses klusterisasi tersebut akan diberikan informasi jumlah iterasi dan waktu proses yang terjadi.

Klusterisasi Algoritma K-Medoid merupakan antarmuka aplikasi klusterisasi rekam medis pasien yang digunakan untuk menjalankan proses pengelompokan data rekam medis pasien menjadi 2 kelompok yaitu Cluster C1 dan Cluster C2. Dalam proses klusterisasi tersebut akan diberikan informasi jumlah iterasi dan waktu proses yang terjadi. Adapun

form klusterisasi algoritma K-Medoid seperti Gambar



Gambar Klusterisasi Algoritma K-Medoid

Output yang menampilkan informasi hasil Perbandingan klusterisasi rekam medis pasien. Adapun output laporan hasil perbandingan klusterisasi, seperti Gambar berikut



Gambar Output Laporan Hasil Perbandingan

Berdasarkan data Rekam Medis, dilakukan klasterisasi menggunakan Metode K-Means dan Metode K-Medoid. Adapun hasil dari klasterisasi tersebut adalah sebagai berikut:

Tabel Hasil Klasterisasi Metode K-Means dan K-Medoid

Metode K-Means			Metode K-Medoid		
Cluster 1			Cluster 1		
Diguna Perspektif	Jumlah Pasien		Diguna Perspektif	Jumlah Pasien	
Diagnosa	1		Diagnosa	1	
K	11		Ekstremal	4	
			CC	11	
Cluster 2			Cluster 2		
Diguna Perspektif	Jumlah Pasien	Gender	Diguna Perspektif	Jumlah Pasien	Gender
Alkoholisitas	2	C	Alkoholisitas	2	C
Alkoholisme	1	C	Alkoholisme	1	C
Asma	1	C	Asma	1	C
CC	1	C	CC	1	C
Diabetes	1	C	Diabetes	1	C
Jumlah Iterasi	4		Jumlah Iterasi	1	
Waktu Proses	2,2955		Waktu Proses	0,0497	

Dari Tabel tersebut hasil dari klasterisasi disimpan. Kemudian membandingkan nilai yang didapatkan dari kedua metode tersebut berdasarkan hasil pengelompokan, jumlah iterasi dan waktu proses yang dibedakan per jenis kelamin.

Adapun hasil perbandingan Metode K-Means dan K-Medoid, seperti Gambar berikut

Perbandingan	Hasil Metode K-Means	Hasil Metode K-Medoid
Hasil Pengelompokan	Cluster 1 = 2 Cluster 2 = 22	Cluster 1 = 3 Cluster 2 = 21
Jumlah Iterasi	4	1
Waktu Proses	2,2955 ms	0,0497 ms

Gambar Hasil Perbandingan Metode K-Means dan K-Medoid

Pada Gambar di atas tersebut dapat disimpulkan bahwa :

1. Berdasarkan hasil pengelompokan dari kedua Metode tersebut, terlihat bahwa terdapat perbedaan pada jumlah cluster yang dihasilkan pada cluster 1 dan cluster 2. Dimana Metode K-Medoid Cluster 1 sebanyak 3 data, dan Cluster 2 sebanyak 21 data, sedangkan Metode K-Means Cluster 1 sebanyak 2 data dan Cluster 2 sebanyak 22 data.
2. Berdasarkan jumlah iterasi, terlihat bahwa Metode K-Medoid hanya membutuhkan iterasi lebih sedikit dibandingkan dengan Metode K-Means. Dimana Metode K-Medoid sebanyak 1 iterasi sedangkan Metode K-Means sebanyak 4 iterasi
3. Berdasarkan waktu proses, terlihat bahwa Metode K-Medoid hanya membutuhkan waktu yang lebih sedikit dibandingkan dengan Metode K-Means. Dimana Metode K-Medoid memiliki waktu 0,0497ms sedangkan Metode K-Means memiliki waktu 2,2955ms, sehingga selisih 2,2458ms lebih cepat Metode K-Medoid.

Dari hasil perbandingan tersebut di atas, maka dapat disimpulkan bahwa Metode K-Medoid lebih baik dari Metode K-Means dari segi jumlah iterasi dan waktu proses yang terjadi karena membutuhkan iterasi lebih sedikit dan waktu yang lebih cepat dalam proses *clustering*.

SIMPULAN

Berdasarkan hasil dari pembahasan dan pengujian yang telah dilakukan, maka dapat disimpulkan bahwa :

1. Pengelompokan data pasien berdasarkan rekam medis pasien dapat digunakan untuk mengetahui penyakit mana yang masuk ke dalam kelompok dengan intensitas tinggi (Cluster C1) atau rendah (Cluster C2).

2. Berdasarkan hasil pengelompokan dari Metode K-Medoid dan Metode K-Means, terlihat bahwa terdapat perbedaan pada jumlah cluster yang dihasilkan pada cluster 1 dan cluster 2.
3. Berdasarkan jumlah iterasi, terlihat bahwa Metode K-Medoid hanya membutuhkan iterasi lebih sedikit dibandingkan dengan Metode K-Means
4. Berdasarkan waktu proses, terlihat bahwa Metode K-Medoid hanya membutuhkan waktu yang lebih sedikit dibandingkan dengan Metode K-Means

DAFTAR PUSTAKA

- Budi, A. 2018. *Pemrograman Windows Dengan Visual Basic .Net : Praktikum Pemrograman VB.Net*. s.l.:Google Book.
- Firman, A., 2019. *Analisis dan Perancangan Sistem Informasi*. Surabaya: Penerbit Qiara Media.
- Indrajani., 2018. *Database Design Theory, Practice, and Case Study*. Jakarta: PT. Elex Media Komputindo.
- Jollyta, D., Ramdhan, W. & Zarlis, M., 2020. *Konsep Data Mining Dan Penerapan*. Yogyakarta: Penerbit Deepublish.
- Kusuma, P. D., 2020. *Machine Learning Teori, Program dan Studi Kasus*. Yogyakarta: Penerbit Deepublish.
- Kusumo, A. S., 2016. *Administrasi SQL Server 2014*. Jakarta: PT. Elex Media Komputindo.
- Lubis, A., 2016. *Basis Data Dasar Untuk Mahasiswa Ilmu Komputer*. Yogyakarta: Deepublish.
- Maulida, L., 2018. Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Provinsi DKI Jakarta Dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, Volume Vol.2 No.3 ISSN:2527-5836.
- Ordila, R., Wahyudi, R., Irawan, Y. & Sari, M. Y., 2020. Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus : Poli Klink PT. Inecda). *Jurnal Ilmu Komputer*, Volume Vol.9 No.2 e-ISSN:2579-3918.
- Prianto, C. & Bunyamin, S., 2020. *Panduan Pembuatan Aplikasi Clustering Gangguan Jaringan Menggunakan Metode K-Means Clustering*. Cetakan Pertama penyunt. Bandung: Penerbit Kreatif Industri Nusantara.
- Suprpto, U., 2021. *Pemodelan Perangkat Lunak (C3) Kompetensi Keahlian : Rekayasa Perangkat Lunak Untuk SMK/MAK Kelas XI*. Jakarta: Grasindo.
- Wahyudi, M., Masitha, Saragih, R. & Solikhun, 2020. *Data Mining : Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering*. Medan: Penerbit Yayasan Kita Menulis.
- Wanto, A. et al., 2020. *Data Mining : Algoritma Dan Implementasi*. Medan: Yayasan Kita Menulis.