

Analisis Perbandingan Algoritma K-Means dan Algoritma K-Medoid Dalam Klasterisasi Penduduk Lanjut Usia

Comparative Analysis of the K-Means Algorithm and the K-Medoid Algorithm in Clustering the Elderly Population

Renal Pujangga¹⁾; Siswanto²⁾; Arius Satoni Kurniawansyah²⁾

^{1,2)}Department of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu Email: ¹⁾

renalpujangga007@gmail.com

How to Cite :

Pujangga, R., Siswanto., Kurniawansyah, A. S. (2021). *Comparative Analysis of the K-Means Algorithm and the K-Medoid Algorithm in Clustering the Elderly Population*. Gatot Kaca Journal , Vol. 2 No. 2 2021 page: 123-128 . DOI: <https://doi.org/10.37638/gatotkaca.2.2.123-128>.

ARTICLE HISTORY

Submitted [29 Desember 2021]

Received [29 Desember 2021]

Revised [30 Desember 2021]

Accepted [31 December 2021]

KEYWORDS

Analysis, Comparison, K-Means Algorithm, K-Medoid Algorithm, Clustering of Elderly Population

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Badan Pusat Statistik merupakan Lembaga Pemerintah Non Kementerian yang bertanggung jawab langsung kepada Presiden. Selama ini data penduduk menurut jenis kelamin dan kelompok umur dilakukan dengan menghitung jumlah penduduk berdasarkan rentang umur yang dimulai dari usia 0-4 tahun sampai dengan 75+ yang dipisah berdasarkan jenis kelamin laki-laki dan perempuan (data terlampir). Data yang selama ini diolah hanya sebatas proyeksi data penduduk, dan tidak diolah kembali untuk mendapatkan informasi tentang pengelompokan data penduduk berdasarkan umur. Maka BPS Kota Bengkulu mengalami kesulitan dalam memetakan program yang terbaik untuk penduduk lanjut usia karena BPS Kota Bengkulu belum memiliki data valid mengenai jumlah penduduk lanjut usia. Pengelompokan data penduduk lanjut usia pada Metode K-Means dan Metode K-Medoids dibagi menjadi 2 kelompok yaitu kelompok tinggi dan rendah. Klasterisasi berdasarkan jenis kelamin Laki-laki diperoleh hasil pengelompokan pada masing-masing clueter berbeda, sedangkan jumlah iterasi yang terjadi sama namun waktu proses antara kedua metode berbeda, dimana metode K-Medoids lebih cepat dibandingkan Metode K-Means. Klasterisasi berdasarkan jenis kelamin Perempuan diperoleh hasil pengelompokan pada masing-masing clueter berbeda, sedangkan jumlah iterasi yang terjadi sama namun waktu proses antara kedua metode berbeda, dimana metode K-Medoids lebih cepat dibandingkan Metode K-Means. 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.

ABSTRACT

The Central Bureau of Statistics is a Non-Ministerial Government Institution that is directly responsible to the President. So far, population data by sex and age group is done by calculating the population based on the age range starting from 0-4 years old up to 75+ which is separated by male and female sex (data attached). The data that has been processed so far is only a projection of population data, and is not reprocessed to obtain information about the grouping of population data by age. So Bengkulu City BPS has difficulty in mapping out the best program for the elderly population because Bengkulu City BPS does not yet have valid data on the number of elderly residents. The grouping of elderly population data in the K-Means Method and the K-Medoids Method was divided into 2 groups, namely high and low groups. Clustering based on male gender, the results of grouping in each clueter are different, while the number of iterations that occur is the same but the processing time between the two methods is different, where the K-Medoids method is faster than the K-Means method. Clustering based on female gender, the results of grouping in each clueter are different, while the number of iterations is the same but the processing time between the two methods is different, where the K-Medoids method is faster than the K-Means method. The results of the comparative analysis between the K-Means and K-Medoids method, it was found that the differences in the results of grouping, iteration and processing time occurred depending on the initial centroid value used in each method.

PENDAHULUAN

Teknologi data mining merupakan salah satu alat bantu untuk penggalian data pada basis data berukuran besar dan dengan spesifikasi tingkat kerumitan yang telah banyak digunakan pada banyak domain aplikasi seperti perbankan, bidang telekomunikasi, dan pemerintahan. Dalam data mining terdapat beberapa metode yang dapat menyelesaikan suatu masalah untuk mendapatkan informasi dari hasil penggalian data, salah satunya yaitu Clustering. Terdapat banyak metode clustering yang dapat diterapkan untuk penggalian data, namun dalam penelitian ini digunakan 2 metode clustering yaitu Metode K-Means dan K-Medoid.

Badan Pusat Statistik merupakan Lembaga Pemerintah Non Kementerian yang bertanggung jawab langsung kepada Presiden. Peranan yang harus dijalankan oleh BPS salah satunya adalah menyediakan kebutuhan data bagi pemerintah dan masyarakat. Data ini didapatkan dari sensus atau survei yang dilakukan sendiri dan juga dari departemen atau lembaga pemerintahan lainnya sebagai data sekunder. Salah satu data yang diolah oleh BPS yaitu data penduduk, dimana data penduduk dibagi menjadi beberapa bagian data penduduk berdasarkan status ekonomi, data penduduk berdasarkan pekerjaan, data penduduk berdasarkan jenis kelamin, data penduduk berdasarkan usia, dan lain-lain. Data-data tersebut diperoleh dari hasil survei atau sensus yang telah dilakukan.

Namun selama ini data penduduk menurut jenis kelamin dan kelompok umur dilakukan dengan menghitung jumlah penduduk berdasarkan rentang umur yang dimulai dari usia 0-4 tahun sampai dengan 75+ yang dipisah berdasarkan jenis kelamin laki-laki dan perempuan (data terlampir). Data yang selama ini diolah hanya sebatas proyeksi data penduduk, dan tidak diolah kembali untuk mendapatkan informasi tentang pengelompokan data penduduk berdasarkan umur. Maka BPS Kota Bengkulu mengalami kesulitan dalam memetakan program yang terbaik untuk penduduk lanjut usia karena BPS Kota Bengkulu belum memiliki data valid mengenai jumlah penduduk lanjut usia.

Oleh karena itu dalam penelitian ini dilakukan kajian terhadap data proyeksi penduduk menurut jenis kelamin dan kelompok umur agar dapat diketahui informasi yang terkandung dari data tersebut berdasarkan jumlah penduduk lanjut usia. Berdasarkan Undang-undang Republik Indonesia Nomor 13 Tahun 1998 Tentang Kesejahteraan Lanjut Usia, Pasal 1 Butir kedua Lanjut usia adalah seseorang yang telah mencapai usia 60 tahun (enam puluh) tahun ke atas.

Selain itu, penulis melakukan kajian terhadap 2 metode clustering tersebut untuk menganalisa perbandingan yang dilihat dari 3 (tiga) aspek yaitu hasil pengelompokan, jumlah iterasi yang terjadi, dan waktu proses pengelompokan. Jumlah klasterisasi yang dilakukan sebanyak 2 cluster yang mengidentifikasi cluster tinggi dan cluster rendah berdasarkan data penduduk lanjut usia selama 11 tahun dari tahun 2010 sampai dengan tahun 2020

LANDASAN TEORI

Clustering

Menurut Prianto (2020:14), Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan atau kemiripan. Clustering juga disebut sebagai segmentasi data dalam beberapa aplikasi. Clustering bersifat unsupervised learning karena tidak membutuhkan proses training dalam pengelompokan kelasnya, tidak seperti klasifikasi yang bersifat supervised learning.

Menurut Jollyta (2020:53), Clustering adalah salah satu model yang dilakukan dengan cara melakukan proses segmentasi terhadap populasi yang heterogen ke dalam sejumlah cluster yang homogen. Proses clustering ini berbeda dengan klasifikasi dimana pada clustering adalah tidak diketahui waktu dimana algoritma dimulai. Clustering dilakukan dengan cara mengumpulkan data yang serumpun dari sebuah data set yang lebih besar. Teknik ini menyingkapkan sejumlah kelompok yang digunakan sebagai masukan datanya.

Algoritma K-Means

Menurut Prianto (2020:18), Algoritma K-Means Clustering merupakan salah satu algoritma dengan partitional, karena K-Means Clustering didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai centroid awalnya. Dibutuhkan jumlah cluster awal yang diinginkan sebagai masukan dan menghasilkan titik centroid akhir sebagai output. Metode K-Means clustering akan memilih pola k sebagai titik awal centroid secara acak atau random. Jumlah iterasi untuk mencapai cluster centroid akan dipengaruhi oleh calon cluster centroid awal secara random. Sehingga didapat cara dalam pengembangan algoritma dengan menentukan centroid cluster yang dilihat dari kepadatan data awal yang tinggi agar mendapatkan kinerja yang lebih tinggi.

Menurut Wahyudi (2020:6), Algoritma K-Means merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat kluster awal. Pada tahap ini pusat kluster dipilih secara acak dari sekumpulan populasi data. Algoritma ini bergantung pada fungsi untuk mengukur data yang mempunyai ciri khas sama. Jarak itu sendiri dihitung menggunakan fungsi euclidean. Kemudian data dimasukkan dalam kelompok yang mempunyai jarak terdekat.

Algoritma K-Medoid

Menurut Wahyudi (2020:9), Metode K-Medoid dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada Tahun 1987. Algoritma K-Medoid sering disebut juga Algoritma Partitioning Around Medoid (PAM). Metode K-Medoid memiliki kesamaan dengan Metode K-Means yaitu sama-sama termasuk Metode Partitioning. Metode Partitioning merupakan metode pengelompokan data dalam sejumlah cluster tanpa adanya struktur hirarki antara satu dengan yang lainnya.

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.

Visual Basic .Net

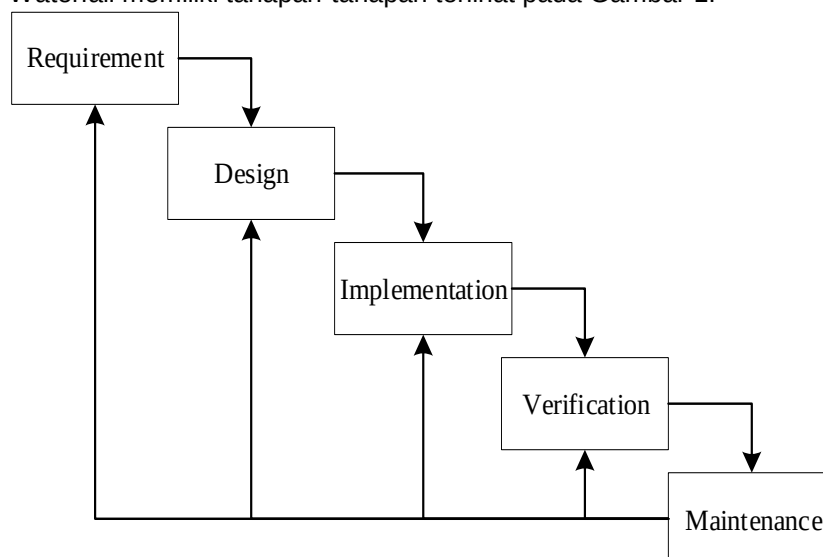
Menurut Blazing (2018:12), Microsoft Visual Basic .Net adalah sebuah alat untuk mengembangkan dan membangun aplikasi yang bergerak di atas sistem .Net Framework, dengan menggunakan bahasa basic. Dengan menggunakan alat ini, para programmer dapat membangun aplikasi windows form, aplikasi web berbasis ASP.Net dan juga aplikasi command-line. Bahasa Visual Basic .Net sendiri menganut paradigma bahasa pemrograman berorientasi objek yang dapat dilihat sebagai evolusi dari Microsoft Visual Basic versi sebelumnya yang diimplementasikan di atas .Net Framework.

Microsoft Visual Studio .Net versi terbaru dengan .Net Framework 4.5 merupakan pengembangan kelanjutan dari Microsoft Visual Studio .Net Framework 4.0 sebelumnya, yaitu Visual Studio .Net 2010 yang diproduksi oleh Microsoft. Pada Bulan Februari Tahun 2002 Microsoft memproduksi teknologi .Net Framework edisi 1.0. Teknologi .Net ini didasarkan atas susunan berupa .Net Framework, sehingga setiap produk baru yang terkait dengan teknologi .Net akan selalu berkembang mengikuti perkembangan .Net Framework dengan berbasis Graphical User Interface (GUI).

METODE PENELITIAN

Metode Analisis

Metode penelitian yang diterapkan pada penelitian ini adalah dengan pengembangan metode waterfall. Metode Waterfall memiliki tahapan-tahapan terlihat pada Gambar 1.

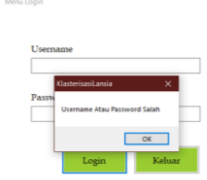
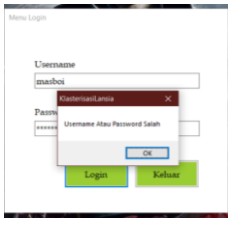
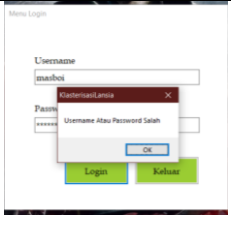
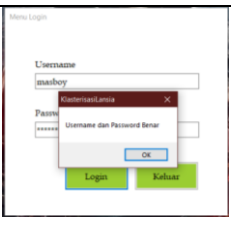


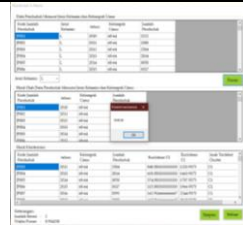
Gambar 1. Tahapan Metode Waterfall

HASIL DAN PEMBAHASAN

Pengujian blackbox (blackbox testing) adalah salah satu metode pengujian perangkat lunak yang berfokus pada sisi fungsionalitas, khususnya pada input dan output Aplikasi Klasterisasi Penduduk Lanjut Usia. Adapun hasil pengujian black box yang telah dilakukan, tampak pada tabel 1.

Tabel 1. Hasil Pengujian Blackbox

Form Uji	Komponen Yang Diuji	Hasil Pengujian Black Box	Kesimpulan
	field username dan password dikosongkan	sistem menolak akses login	
	mengisi username yang salah	sistem menolak akses login	
	mengisi password yang salah	sistem menolak akses login	
	mengisi username dan password yang benar	sistem menerima akses login	
Input Data Penduduk	menyimpan data penduduk yang berbeda	sistem berhasil menyimpan data tersebut	

	menyimpan data penduduk yang sama	sistem menolak akses simpan data tersebut	
Klasterisasi Algoritma K-Means	melakukan proses klasterisasi data penduduk lanjut usia	sistem berhasil menampilkan hasil klasterisasi dan memberikan informasi jumlah iterasi, waktu proses dan hasil pengelompokan.	
Klasterisasi Algoritma K-Medoid	melakukan proses klasterisasi data penduduk lanjut usia	sistem berhasil menampilkan hasil klasterisasi dan memberikan informasi jumlah iterasi, waktu proses dan hasil pengelompokan.	

Berdasarkan hasil pengujian yang telah dilakukan, Aplikasi Klasterisasi Penduduk Lanjut Usia berhasil dilakukan, dan dapat memberikan informasi berdasarkan 2 kelompok yaitu Cluster C1 (Tinggi) dan Cluster C2 (Rendah), serta fungsional dari aplikasi telah berjalan sesuai dengan yang diharapkan.

KESIMPULAN DAN SARAN

Kesimpulan

1. Pengelompokan data penduduk lanjut usia pada Metode K-Means dan Metode K-Medoids dibagi menjadi 2 kelompok yaitu kelompok tinggi dan rendah.
2. Klasterisasi berdasarkan jenis kelamin Laki-laki diperoleh hasil pengelompokan pada masing-masing cluster berbeda, sedangkan jumlah iterasi yang terjadi sama namun waktu proses antara kedua metode berbeda, dimana metode K-Medoids lebih cepat dibandingkan Metode K-Means
3. Klasterisasi berdasarkan jenis kelamin Perempuan diperoleh hasil pengelompokan pada masing-masing cluster berbeda, sedangkan jumlah iterasi yang terjadi sama namun waktu proses antara kedua metode berbeda, dimana metode K-Medoids lebih cepat dibandingkan Metode K-Means
4. 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

Saran

Penulis menyarankan perlu dilakukan pengembangan pada penelitian selanjutnya dengan menambahkan algoritma clustering yang lainnya sebagai bahan perbandingan.

DAFTAR PUSTAKA

- Blazing, 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.
- 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.