

ANALISIS PENCATATAN AKTA KELAHIRAN MENGGUNAKAN METODE CLUSTERING

Reza Revo Fahlefi¹, Zunaida Sitorus²

Universitas Asahan

e-mail: ¹rezarevo16@gmail.com, ²z_sitorus@yahoo.com

Abstract: *This study explores the application of the K-Means Clustering algorithm in grouping birth certificate data in Labuhanbatu Utara Regency. The aim of this research is to identify patterns in birth certificate ownership across various regions and to facilitate data clustering to support government planning related to birth certificate ownership. The analysis reveals that the K-Means Clustering method is effective in categorizing data according to the level of birth certificate ownership. Regions such as Kualuh Hulu, NA IX-X, and Kualuh Selatan have high levels of birth certificate ownership, while Kualuh Hilir, Aek Kuo, and Marbau fall into the medium category. Meanwhile, Kualuh Leidong and Aek Natas show low levels of birth certificate ownership. This information is crucial for designing targeted outreach programs to raise public awareness of the importance of having a birth certificate. The consistency of clustering results between the first and second iterations demonstrates the stability and efficiency of the K-Means method, achieving accurate outcomes without the need for multiple iterations. Therefore, the K-Means algorithm proves to be a reliable data analysis tool for decision-making in government programs.*

Keywords: *K-Means Clustering, Birth Certificate, Data Clustering.*

Abstrak: Penelitian ini membahas penerapan algoritma K-Means Clustering dalam pengelompokan data akta kelahiran di Kabupaten Labuhanbatu Utara. Tujuan dari penelitian ini adalah untuk mengidentifikasi pola kepemilikan akta kelahiran di berbagai wilayah serta mempermudah proses pengelompokan data guna mendukung perencanaan program pemerintah terkait kepemilikan akta kelahiran. Penelitian ini menggunakan data kuantitatif dari Disdukcapil Kabupaten Labuhanbatu Utara. Berdasarkan analisis yang dilakukan, hasil menunjukkan bahwa metode K-Means Clustering efektif dalam mengelompokkan data sesuai dengan tingkat kepemilikan akta kelahiran. Wilayah Kualuh Hulu, NA IX -X, dan Kualuh Selatan memiliki tingkat kepemilikan yang tinggi, sementara Kualuh Hilir, Aek Kuo, dan Marbau berada pada kategori sedang. Sedangkan wilayah Kualuh Leidong dan Aek Natas menunjukkan tingkat kepemilikan akta kelahiran yang rendah. Informasi ini penting untuk merancang program penyuluhan yang lebih tepat sasaran guna meningkatkan kesadaran masyarakat akan pentingnya memiliki akta kelahiran. Konsistensi hasil clustering antara iterasi pertama dan kedua menunjukkan stabilitas dan efisiensi metode K-Means, dengan hasil yang akurat dapat dicapai tanpa memerlukan banyak iterasi. Dengan demikian, algoritma K-Means dapat diandalkan sebagai alat analisis data dalam pengambilan keputusan untuk program pemerintah.

Kata kunci: K-Means Clustering, Akta Kelahiran, Pengelompokan Data.

PENDAHULUAN

Dinas Kependudukan dan Pencatatan Sipil Kabupaten Labuhanbatu Utara (Disdukcapil Labura) adalah bagian penting dari administrasi pemerintahan

yang bertanggung jawab atas berbagai urusan terkait dengan penduduk, mulai dari pembuatan dan pemeliharaan data kependudukan hingga pelayanan terkait status sipil seperti kelahiran, kematian, perkawinan, dan perceraian. Kelahiran

merupakan salah satu peristiwa penting dalam hidup. Semua kelahiran dicatat dengan alasan bahwa akan mempengaruhi pendidikan dan membawa jaminan yang sah kepada orang-orang yang bersangkutan dan daerah sekitarnya. Salah satu indikator keberhasilan Dinas Kependudukan dan Pencatatan Sipil Kabupaten Labuhanbatu Utara adalah terpenuhinya kepemilikan akta kelahiran pada setiap individu. Permasalahan yang dihadapi Dinas Kependudukan dan Pencatatan Sipil Kabupaten Labuhanbatu Utara terkait kepemilikan akta kelahiran adalah tidak sebandingnya jumlah kepemilikan akta kelahiran dengan jumlah penduduk di Kabupaten Labuhanbatu Utara. Hal ini disebabkan kurangnya pemetaan dan pengelompokan data kepemilikan akta kelahiran.

Untuk itu pengelompokan data akta kelahiran pada setiap Kecamatan yang ada memerlukan penerapan metode Data Mining. Data Mining atau yang sering disebut dengan istilah *Knowledge discovery in Database* (KDD) merupakan sebuah proses untuk menemukan suatu pola dari kumpulan data. Data Mining memiliki peranan yang sangat penting dalam pengolahan data berskala besar di dalam beberapa bidang kehidupan di antaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu dan teknologi. *Clustering* adalah metode untuk membagi sekumpulan titik data ke dalam kelompok atau *Cluster* tertentu. Cara kerja *Clustering* adalah dengan mencari data dan mengumpulkan informasi yang memiliki sifat komparatif antara informasi yang diambil dengan informasi yang berbeda. Dalam teknik pengelompokan, terdapat dua jenis analisis *Cluster* yang menggunakan algoritma terkait yaitu K-means dan K-medoids. Dalam penelitian ini, peneliti menggunakan algoritma K-means dari teknik *Clustering* untuk membantu mengelompokkan data.

Penelitian ini menerapkan teknik algoritma K-means dalam mengelompokkan data akta kelahiran. Diharapkan dengan penerapan algoritma

tersebut kita dapat mengelompokkan data kepemilikan akta kelahiran sehingga dapat sebanding dengan jumlah penduduk di kabupaten Labuhanbatu Utara.

Berdasarkan permasalahan data di atas maka penulis tertarik mengambil judul “Analisis Pencatatan Akta Kelahiran Menggunakan Metode Clustering”

METODE

Penelitian ini menggunakan data kuantitatif dari Disdukcapil Kabupaten Labuhanbatu Utara. Pengumpulan data yang digunakan untuk penelitian terdapat beberapa metode yang terdiri dari:

1. Penelitian Kepustakaan (*Library Research*) yaitu memanfaatkan perpustakaan, buku, prosiding atau journal sebagai media untuk bahan referensi dalam menentukan faktor, parameter dan label yang digunakan untuk penelitian.
2. Penelitian Lapangan (*Field Work Research*) yaitu penelitian yang dilakukan secara langsung ke kantor Disdukcapil. Penulis memperoleh informasi dengan mengumpulkan data, mempelajari data, validasi data dan mencari referensi terkait dengan kasus pada penelitian. Luaran dari studi literature ini adalah tersusun dan terkoleksinya referensi yang baik dan benar dengan penelitian.

Seleksi Data (Data Selection)

Seleksi data merupakan proses memilih kumpulan data, atau berfokus pada subset variabel atau sampel data. Oleh karena itu hanya beberapa data yang akan diambil. Pada tahap seleksi ini akan dilakukan proses seleksi atribut, seperti atribut mana yang digunakan oleh peneliti pada data mining, pada data jumlah kepemilikan akta kelahiran ini terdapat 6 atribut awal yaitu, nama provinsi, kode wilayah, nama wilayah, tahun, jumlah penduduk pemilik akta kelahiran, jumlah penduduk yang akan diseleksi menjadi 3 atribut, nama wilayah, jumlah penduduk pemilik akta kelahiran serta tahun.

Pembersihan Data (Data Cleaning)

Pembersihan data perlu dilakukan untuk menghilangkan kebisingan atau data yang tidak konsisten, mengumpulkan informasi yang diperlukan untuk memperhitungkan kebisingan, memutuskan strategi untuk menangani data yang hilang. Pada penelitian ini tidak ditemukan data duplikat ataupun data kosong.

HASIL DAN PEMBAHASAN

Dari penelitian yang telah dilakukan maka selanjutnya adalah tahapan proses K-means. Untuk lebih jelasnya dapat dilihat pada tabel berikut :

Analisis Metode K-Means

Setelah diperoleh data akta kelahiran, maka selanjutnya adalah menentukan rentang kluster. Untuk menentukan rentang kluster yang tepat, beberapa pendekatan dan rentang jumlah kluster dapat digunakan:

1. Minimum K: Minimal kita bisa memilih $K = 2$, karena jika hanya satu kluster ($K = 1$), semua data akan berada dalam satu kelompok, yang tidak memberikan informasi yang berguna dalam klasterisasi.
2. Maksimum K: Dengan jumlah data 8 wilayah, jumlah kluster maksimum yang mungkin secara teori adalah $K = 8$ (dimana setiap wilayah membentuk kluster tersendiri). Namun, ini tidak memiliki makna yang jelas karena tujuan klasterisasi adalah mengelompokkan data yang mirip.
3. Rentang Kluster yang Wajar: Berdasarkan ukuran data ini, rentang jumlah kluster yang wajar dan praktis dapat dipertimbangkan antara $K = 2$ hingga $K = 4$. Ini karena kita ingin:
 - a. $K = 2$ jika kita hanya ingin memisahkan wilayah dengan nilai tinggi dan rendah.
 - b. $K = 3$ untuk mendapatkan kelompok rendah, sedang, dan tinggi (yang sesuai dengan data).
 - c. $K = 4$ jika ada pola lebih kompleks di dalam data (misalnya satu kluster tambahan untuk wilayah dengan nilai yang sangat berbeda).

Dari beberapa pendekatan dan rentang jumlah kluster dapat digunakan, maka peneliti memutuskan untuk menggunakan kluster sebanyak 3, sesuai dengan pengelompokan yaitu tinggi, sedang, dan rendah.

Selanjutnya adalah melakukan perhitungan pada data akta kelahiran untuk mendapat 3 data nilai centroid berdasarkan nilai rata-rata pada data akta kelahiran.

Pemilihan Centroid Awal

Centroid Awal 1: Pilih nilai rata-rata dari tiga wilayah pertama: Kualuh Hulu, Kualuh Leidong, dan Kualuh Hilir.

$$X_{C1} = \frac{1997 + 870 + 925}{3} = 1264$$

$$Y_{C1} = \frac{2071 + 1183 + 834}{3} = 1362,67$$

Centroid Awal 2: Pilih nilai rata-rata dari tiga wilayah pertama: Kualuh Hulu, Kualuh Leidong, dan Kualuh Hilir.

$$X_{C2} = \frac{984 + 1438 + 2004}{3} = 1475,33$$

$$Y_{C2} = \frac{837 + 1023 + 1435}{3} = 1098,33$$

Centroid Awal 3: Pilih nilai rata-rata dari tiga wilayah pertama: Kualuh Hulu, Kualuh Leidong, dan Kualuh Hilir.

$$X_{C3} = \frac{1163 + 1897}{3} = 1530$$

$$Y_{C3} = \frac{1217 + 1982}{3} = 1599,5$$

Tabel 1. Data Nilai Centroid

No	Centroid	Tahun 2022	Tahun 2023
1	Centroid 1	1264	1362,67
2	Centroid 2	1475,33	1098,33
3	Centroid 3	1530	1599,5

Proses perhitungan *K-Means* dimulai dari perulangan 1 sampai dengan nilai *cluster* sebelum sama dengan nilai *cluster* pada perulangan terakhir, semua proses akan dijabarkan termasuk perhitungan jarak terdekat yaitu *euclidean distance* menggunakan rumus:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

Dimana:

x_1 dan y_1 adalah koordinat data yang diberikan (dalam hal ini populasi di tahun 2022 dan 2023).

x_2 dan y_2 adalah koordinat centroid awal yang diberikan.

Berikut tabel dari hasil perhitungan *Euclidean Distance*:

Tabel 2. Data Perhitungan *Euclidean Distance*

No	Wilayah	C1	C2	C3	Cluster
1	Kualuh Hulu	1019.33	1103.73	663.63	3
2	Kualuh Leidong	433.03	611.23	780.43	1
3	Kualuh Hilir	628.02	610.52	975.71	2
4	Aek Kuo	595.59	556.51	937.83	2
5	Marbau	381.64	84.08	583.79	2
6	NA IX – X	743.53	626.76	501.73	3
7	Aek Natas	177.26	334.12	530.09	1
8	Kualuh Selatan	885.59	979.12	530.09	3

Setelah perhitungan pertama selesai maka akan dilanjutkan perhitungan kedua dengan melakukan penentuan nilai centroid baru.

Pemilihan Centroid Perulangan 2

Centroid 1: terdapat 2 wilayah yaitu: Kualuh Leidong, dan Aek Natas.

$$X_{C1} = \frac{870 + 1163}{2} = 1016,5$$

$$Y_{C1} = \frac{1183 + 1217}{2} = 1200$$

Centroid 2: terdapat 3 wilayah yaitu: Kualuh Hilir, Aek Kuo dan Marbau.

$$X_{C2} = \frac{925 + 984 + 1438}{3} = 1115,66$$

$$Y_{C2} = \frac{834 + 837 + 1023}{3} = 898$$

Centroid 3: terdapat 3 wilayah yaitu: Kualuh Hulu, NA IX-X dan Kualuh Selatan

$$X_{C3} = \frac{1997 + 2004 + 1897}{3} = 1966$$

$$Y_{C3} = \frac{2071 + 1435 + 1982}{3} = 1829,33$$

Tabel 3. Data Nilai Centroid Baru

Centroid	2022	2023
Centroid 1	1016,5	1200
Centroid 2	1115,66	898
Centroid 3	1966	1829,33

Setelah Centroid baru didapatkan maka akan dilakukan perhitungan *euclidean distance* kembali

Tabel 4. Data Perhitungan *Euclidean Distance* ke 2

No	Wilayah	C1	C2	C3	Cluster
1	Kualuh Hulu	1311.50	1467.20	243.65	3
2	Kualuh Leidong	147.48	376.27	1272.38	1
3	Kualuh Hilir	377.26	201.12	1440.27	2
4	Aek Kuo	364.45	145.11	1396.08	2
5	Marbau	457.16	345.72	963.82	2
6	NA IX -X	1015.08	1038.03	396.16	3
7	Aek Natas	147.48	322.49	1009.83	1
8	Kualuh Selatan	1177.63	1336.24	167.54	3

Berdasarkan perhitungan *euclidean distance* pertama dan kedua, data *cluster* sama yang berarti perhitungan berhenti sampai perhitungan kedua saja, dan dari

data perhitungan tersebut maka dianalisis sebagai berikut.

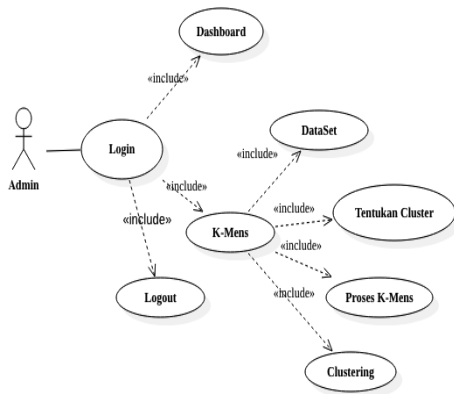
1. Kualuh Hulu, NA IX -X, dan Kualuh Selatan kepemilikan akta kelahiran tinggi
2. Kualuh Hilir, Aek Kuo, dan Marbau kepemilikan akta kelahiran sedang, Kualuh Leideong dan Aek Natas kepemilikan akta kelahiran rendah.

Perancangan Sistem

Perancangan sistem menguraikan bagaimana alur proses *input* maupun *output* dari sistem yang akan dihasilkan. Perancangan sistem ini dapat digambarkan melalui diagram UML (*Unified Modelling Language*) yang akan menggambarkan aliran data terhadap sistem yang dirancang.

Use Case Diagram

Untuk memperjelas bentuk rancangan sistem dan mempermudah dalam menganalisa sistem yang dirancang, harus ada gambaran yang lebih *detail* tentang elemen-elemen dan hubungannya serta pembahasan dari setiap ruang lingkup sistem. Salah satu media yang dapat digunakan untuk memperjelas hal tersebut adalah *use case diagram*. *Use case diagram* merupakan fungsionalitas dari suatu sistem yang dapat dilihat pada gambar 1.

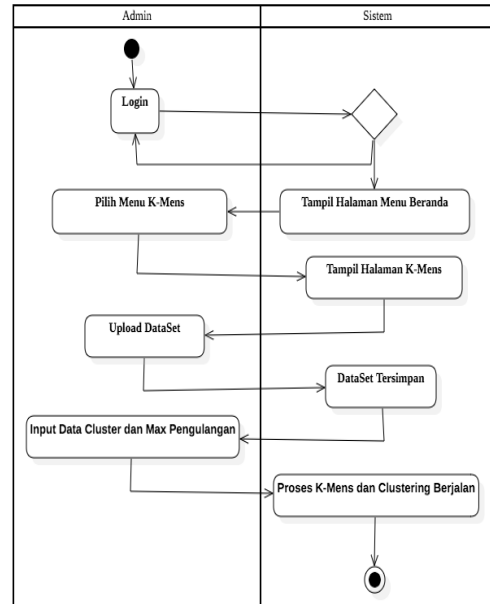


Gambar 1. Use Case Diagram

Activity Diagram

Activity diagram merupakan salah satu cara memodelkan *event-event* yang terjadi dalam *use case*. Pada diagram ini secara esensial mirip dengan diagram alir (*flowchart*), memperlihatkan aliran kendali dari suatu aktivitas ke aktivitas

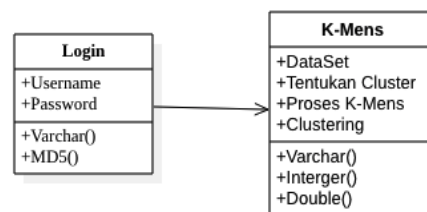
lainnya. *Activity diagram* berfungsi untuk memvisualisasikan, menspesifikasi, serta mendokumentasikan sifat dari sekumpulan objek, selain itu juga dapat digunakan memodelkan aliran kendali dari suatu operasi. Untuk lebih jelasnya dapat dilihat pada gambar berikut ini:



Gambar 2. Activity Diagram

Class Diagram

Class Diagram merupakan bentuk bagan yang menggunakan relasi dan entitas yang dibuat dengan menggunakan persepsi yang terdiri dari sekumpulan objek dasar yaitu entitas dan hubungan antar entitas.



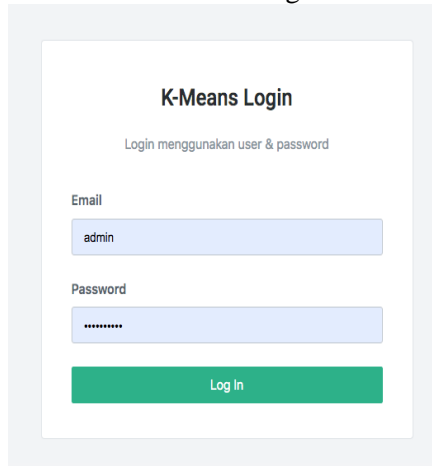
Gambar 3. Class Diagram

Implementasi Sistem

Tahap implementasi sistem merupakan tahapan penerapan sistem agar siap untuk dioperasikan. Tujuan implementasi sistem adalah untuk menjelaskan manual modul kepada *user* yang akan menggunakan sistem, sehingga *user* dapat merespon apa yang ditampilkan oleh sistem.

Tampilan Menu *Login*

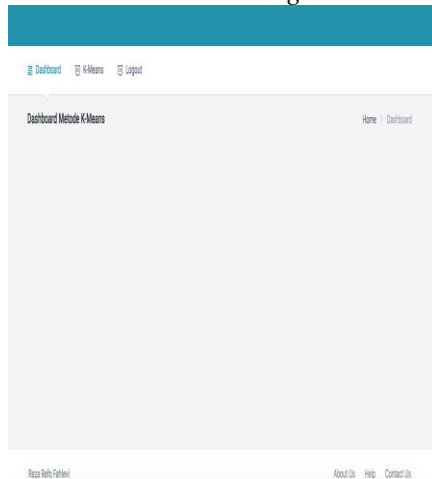
Pada menu login terdapat inputan username dan password sebagai validasi data untuk dapat mengakses inputan sistem. Berikut merupakan tampilan menu login pada sistem analisis data akta kelahiran metode clustering.



Gambar 4. Tampilan Menu *Login*

Tampilan Menu Beranda

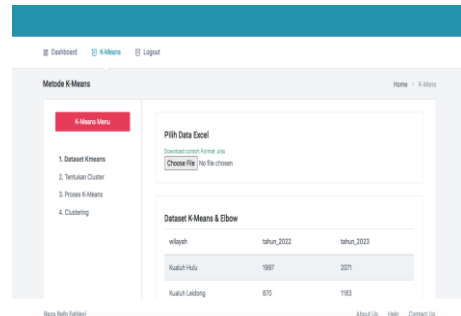
Menu beranda merupakan tampilan awal saat login berhasil dilakukan. Berikut merupakan tampilan menu beranda pada sistem analisis data akta kelahiran metode *clustering*.



Gambar 5. Tampilan Menu Beranda

Tampilan Menu *DataSet*

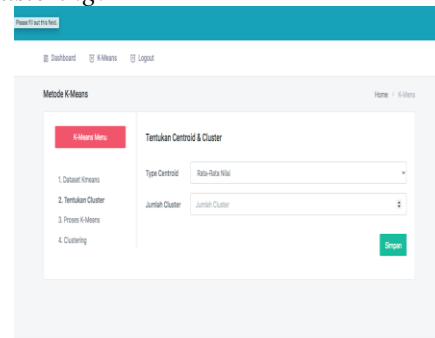
Menu dataset merupakan menu untuk melakukan upload data akta kelahiran ke sistem. Berikut merupakan tampilan menu dataset pada sistem analisis data akta kelahiran metode *clustering*.



Gambar 6. Tampilan Menu *Data Set*

Tampilan Menu *Data Centroid dan Cluster*

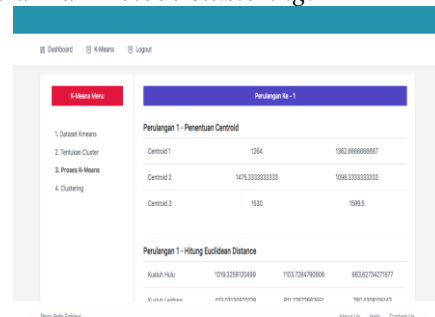
Menu data centroid dan cluster merupakan menu untuk melakukan input jumlah centroid dan cluster yang digunakan. Berikut merupakan tampilan menu data centroid dan cluster pada sistem analisis data akta kelahiran metode *clustering*.



Gambar 7. Tampilan Menu *Data Centroid dan Cluster*

Tampilan Menu *Proses K-Means*

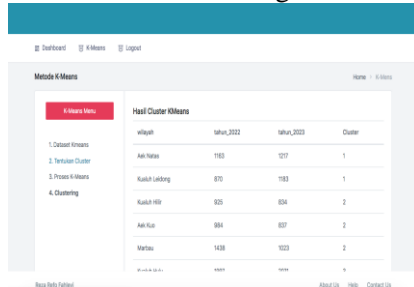
Menu proses K-Means merupakan menu untuk melihat hasil dari perhitungan *euclidean distance* yang dilakukan. Berikut merupakan tampilan menu proses K-means pada sistem analisis data akta kelahiran metode *clustering*.



Gambar 8. Tampilan Menu *Proses K-Means*

Tampilan Menu Klustering

Menu Klustering merupakan menu untuk melihat hasil dari perhitungan. Berikut merupakan tampilan menu klustering pada sistem analisis data akta kelahiran metode clustering.



The screenshot shows a web application interface for K-Means clustering. It features a sidebar with navigation options: 'K-Means Info', '1. Cluster K-Means', '2. Visualisasi Cluster', '3. Proses K-Means', and '4. Clustering'. The main content area displays a table titled 'Hasil Cluster K-Means' with columns for 'Wilayah', 'tahun_2022', 'tahun_2023', and 'Cluster'. The table contains the following data:

Wilayah	tahun_2022	tahun_2023	Cluster
Aek Natas	1183	1217	1
Kualuh Lading	870	1183	1
Kualuh Hilir	825	824	2
Aek Kuo	984	827	2
Marbau	1428	1023	2

Gambar 9. Tampilan Menu Klustering

SIMPULAN

Berdasarkan penelitian, implementasi dan pengujian, maka dapat diambil kesimpulan sebagai berikut:

1. Penerapan algoritma *K-Means Clustering* dalam pengelompokan data akta kelahiran di Kabupaten Labuhanbatu Utara menunjukkan bahwa metode ini efektif dalam mengidentifikasi pola kepemilikan akta kelahiran di berbagai wilayah yang dijadikan bahan penelitian.
2. Hasil analisis menunjukkan bahwa wilayah Kualuh Hulu, NA IX -X, dan Kualuh Selatan memiliki tingkat kepemilikan akta kelahiran yang tinggi, sedangkan Kualuh Hilir, Aek Kuo, dan Marbau menunjukkan kepemilikan sedang. Wilayah Kualuh Leidong dan Aek Natas memiliki tingkat kepemilikan akta kelahiran yang rendah. Hal ini dapat dijadikan sebagai informasi penting untuk perencanaan dan pengembangan program penyuluhan mengenai pentingnya akta kelahiran.
3. Konsistensi hasil *cluster* antara perhitungan *Euclidean distance* pertama dan kedua menunjukkan stabilitas dalam pengelompokan data, yang berarti bahwa algoritma *K-Means* berhasil mengidentifikasi *cluster* dengan baik, sehingga perhitungan dapat dihentikan setelah

iterasi kedua. Ini menunjukkan efisiensi metode dalam mencapai hasil yang konsisten tanpa memerlukan banyak iterasi.

DAFTAR PUSTAKA

- Achray, Aditya Agusti. 2020. Implementasi Algoritma K-Means Untuk Mengelompokkan Data Penjualan Mobil Di Pt . Honda Arista Mangga Dua Implementasi Algoritma K-Means Untuk Mengelompokkan Data Penjualan Mobil Di Pt .
- Ananda, Dhea, Siti Rohimah, Bagas Susilo, Damar Wulan, And Asdar Mustofa. 2022. "Implementasi K-Means Dalam Pengelompokkan Data Akta Kelahiran Di Indonesia: Implementation Of K-Means In Grouping Birth Certificate Data In Indonesia." Sentimas: Seminar Nasional Penelitian Dan Pengabdian Masyarakat 66–71.
- Achray, Aditya Agusti. 2020. Implementasi Algoritma K-Means Untuk Mengelompokkan Data Penjualan Mobil Di Pt . Honda Arista Mangga Dua Implementasi Algoritma K-Means Untuk Mengelompokkan Data Penjualan Mobil Di Pt .
- Ananda, Dhea, Siti Rohimah, Bagas Susilo, Damar Wulan, And Asdar Mustofa. 2022. "Implementasi K-Means Dalam Pengelompokkan Data Akta Kelahiran Di Indonesia: Implementation Of K-Means In Grouping Birth Certificate Data In Indonesia." Sentimas: Seminar Nasional Penelitian Dan Pengabdian Masyarakat 66–71.
- Event, And Intan Utnasari. 2021. "Analisis Clustering Dengan K-Means Untuk Pengelompokkan Penjualan Produk Pada Hotel Newton." Computer And Science Industrial Engineering 04(Vol 4 No 4 (2021): Comasie):1–8. [4] A. F. Baba D. Kuşçu, and K. Han,

- “Developing a Software for Fuzzy Group Decision Support System: a Case Study.” *Turkish Online J. Educ. Technol.*, vol. 8, no. 3, pp. 22–29, 2009.
- Hilmi, Rafiqi Zul, Ratih Hurriyati, And Lisnawati. 2018. “Penerapan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Universitas Islam Negeri Sunan Kalijaga Yogyakarta Skripsi.” 3(2):91–102.
- Khaira, Ulfa, S. Komp, And M. Kom. 2023. *Penambangan Data Dan Intelengensi Bisnis*. Vol. 1.
- Napitupulu, Flora Sabarina, Irfan Sudahri Damanik, Ilham Syahputra Saragih, And Anjar Wanto. 2020. “Algoritma K-Means Untuk Pengelompokan Dokumen Akta Kelahiran Pada Tiap Kecamatan Di Kabupaten Simalungun.” *Building Of Informatics, Technology And Science (Bits)* 2(1):55–63.
- Rahmayanti, Diah Ayu. 2022. “Penerapan Metode K-Means Untuk Clustering Data Anak Berdasarkan Kepemilikan Akta Kelahiran Dan Kia.” *Informal: Informatics Journal* 7(3):210.
- Saputra, William, Sundari Retno Andani, Indra Gunawan, Zulaini Masruro Nasution, Teknik Informatika, Stikom Tunas Bangsa Pematangsiantar, And Jl A. Jend Sudirman Blok No. 2022. “Penerapan Data Mining Dalam Menentukan Standard Kelayakan Penerima Insentif Pegawai Pt. Sinarmas Multifinance Menggunakan Algoritma K-Means.” *Maret* 7:10–17.
- Wahyuningsih, Listiyani, And Ahmad Rifai. 2023. “Implementasi Algoritma K-Means Terhadap Pengelompokan Kepemilikan Akta Kelahiran Di Jawa Barat.” *E-Link: Jurnal Teknik Elektro Dan Informatika* 18(1):1.