

Pemanfaatan Data Rekam Medis dalam Menentukan Cluster Penyakit Melalui Data Mining di Rs Aisyiyah Siti Fatimah Tulangan Sidoarjo

Brema Arya Yudano^{1a}, Uce Indahyanti^{2b}, Umi Khoirun Nisak^{3c}*

^{1,2,3} Universitas Muhammadiyah Sidoarjo, Jl Raya Rame Pilang No, Sidoarjo 61261, Indonesia

^a bremaarya@umsida.ac.id*; ^b uceindahyanti@umsida.ac.id; ^c umikhoirun@umsida.ac.id

* Co-Author

INFORMASI ARTIKEL	ABSTRAK
Sejarah artikel: Tanggal diterima : 30 Oktober 2022 Tanggal revisi : 01 Nopember 2022 Diterima : 28 Nopember 2022 Diterbitkan : 23 Desember 2022	Jumlah rekam medis yang masuk setiap hari mengalami penambahan jumlah data di rumah sakit Aisyiyah Siti Fatimah Tulangan Sidoarjo. Namun, data yang tersimpan terbatas menjadi tumpukan data yang belum tergalai secara optimal dalam pemanfaatannya. Pemanfaatan data yang belum optimal seharusnya dapat dimanfaatkan untuk menghasilkan informasi baru jika digali lebih dalam dengan metode data mining. Penelitian ini bertujuan melakukan clustering diagnosa penyakit berdasarkan data rekam medis menggunakan algoritma K-Means. Metode penelitian menggunakan deskriptif. Hasil proses clustering software RapidMiner dengan 5 cluster mendapatkan hasil C ₀ didominasi diagnosa I11.0, C ₁ didominasi diagnosa E11.8, C ₂ didominasi diagnosa A15.3, C ₃ didominasi diagnosa E05.9, C ₄ didominasi diagnosa E11.8. Uji validitas cluster menggunakan metode <i>Davies Bouldin Index</i> (DBI) dengan 5 cluster mendapatkan nilai 0.469
Kata Kunci : Rekam Medis Cluster Penyakit Data Mining	

Copyright (c) 2022 Care Journal

This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

PENDAHULUAN

Rumah sakit merupakan sarana kesehatan untuk masyarakat yang mempunyai tujuan untuk menyelenggarakan pelayanan kesehatan tiap orang secara paripurna. Rumah sakit memiliki tugas dan fungsi memberikan pelayanan kesehatan perorangan baik pelayanan rawat darurat, rawat jalan, rawat inap, pelayanan rujukan yang dapat mencakup pelayanan dalam rekam medis dan penunjang medis serta dapat dimanfaatkan juga sebagai bahan pendidikan, pelatihan, dan penelitian (Kemenkes RI, 2009).

Rekam medis merupakan bukti catatan dan dokumen tentang identitas pasien, pemeriksaan, pengobatan, tindakan dan pelayanan lain yang telah diberikan kepada pasien (Kemenkes RI, 2008). Dengan adanya bukti tertulis dalam rekam medis. Maka, data rekam medis dapat dipakai sebagai pertanggung jawaban dengan maksud dan tujuan sebagai penunjang tertib administrasi sebagai upaya dalam optimalisasi pelayanan Kesehatan (Wijaya & Vera, 2021).

Clustering merupakan metode analisis data yang sering dimasukkan ke dalam salah satu metode *data mining* yang bertujuan untuk melakukan pengelompokan data dengan karakteristik yang sama ke suatu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain (Mochamad Wahyudi, dkk, n.d.). *Clustering* berbeda dengan *group*, *group* merupakan suatu kelompok yang mempunyai kondisi yang sama sedangkan *cluster* tidak harus selalu sama melainkan pengelompokannya didasarkan pada kedekatan dari satu karakteristik sampel yang ada dan salah satunya dengan menggunakan rumus jarak *euclidean* (Amir Ali, n.d.).

Berdasarkan beberapa studi pendahuluan penelitian terdahulu terkait dengan *clustering* data rekam medis, dimana salah satu penelitian yang telah dilakukan

menghasilkan data berupa 3 (tiga) kelompok *cluster* yaitu: wilayah, jenis kelamin, dan umur dengan jumlah pasien berdasarkan kelompok *cluster* usia dewasa 4912 pasien, *cluster* kedua adalah anak-anak 1262 pasien, dan *cluster* ketiga adalah balita 144 pasien. Dengan jumlah penyakit endemik terbanyak yaitu infeksi saluran pernapasan atas (ISPA) dengan jumlah 1985 pasien dan juga penyakit lain seperti jatuh dari motor, cek kolesterol, kontrol kehamilan, cek tensi, dll dengan jumlah 2142 pasien (Ordila et al., 2020).

Penelitian lain menggunakan algoritma *K-Means* yang disusun oleh (Rahmayani, 2018) [9] dengan sampel data sebanyak 278 pasien dengan hasil keseluruhan menunjukkan tingkat keparahan penyakit pasien *cluster* (C_0) berada pada penyakit berat dengan jumlah 47 pasien, *cluster* (C_1) berada pada penyakit ringan dengan jumlah 82 pasien, dan *cluster* (C_2) berada pada penyakit sedang dengan jumlah 149 pasien. Dengan tingkat keparahan penyakit yang diderita pasien berada pada penyakit sedang dengan persentase 53,59%, penyakit paling sering diderita masyarakat sekitar adalah ISPA.

Berdasarkan uraian latar belakang diatas maka penulis mengusulkan penelitian mengenai *implementasi* teknologi *data mining* untuk *clustering* data rekam medis di Rumah Sakit 'Aisyiyah Siti Fatimah Tulangan Sidoarjo berdasarkan diagnosis penyakit dengan menggunakan *attribute* umur, jenis kelamin, wilayah, dan diagnosis ICD-10 menggunakan algoritma *K-Means* dengan alat bantu perhitungan berupa *software* RapidMiner. Sedangkan hasil yang didapatkan dari metode *clustering* ini diharapkan dapat dijadikan sebagai suatu acuan bagi pihak rumah sakit untuk mengambil suatu kebijakan dalam melakukan tindakan *preventif* dengan memberikan penyuluhan kesehatan terhadap wilayah yang banyak terjangkau penyakit di kabupaten Sidoarjo.

BAHAN DAN METODE

Metode penelitian ini menggunakan jenis penelitian deskriptif. Penelitian deskriptif merupakan penelitian yang digunakan untuk menggambarkan gejala-gejala, fakta-fakta atau kejadian-kejadian secara sistematis dan akurat, mengenai sifat-sifat populasi atau daerah tertentu (Adhi Kusumastuti et al., 2020). Penelitian ini berlokasi di Rumah Sakit 'Aisyiyah Siti Fatimah Tulangan Sidoarjo pada tahun 2021. Sumber data dalam penelitian ini berasal dari data rekam medis laporan triwulan kunjungan pasien rawat jalan dan rawat inap pada bulan Agustus-Oktober tahun 2021 yang diperoleh secara langsung dari proses observasi dan wawancara. Populasi dan sampel dalam penelitian ini dihitung menggunakan rumus slovin dengan hasil yang didapatkan sebesar 369 data. Proses analisa data dilakukan dengan beberapa tahapan berupa *preprocessing* data. *Preprocessing* data merupakan metode dari *data mining* untuk menghasilkan data yang berkualitas (Mustika Mustika et al., 2021). Metode pengolahan data dalam penelitian ini menggunakan *software* RapidMiner sebagai alat bantu untuk menerapkan *algoritma K-Means clustering*.

Clustering adalah salah satu alat bantu dalam melakukan proses *data mining* yang bertujuan untuk pengelompokan objek kedalam *cluster* sehingga data yang ada dalam satu *cluster* mempunyai tingkat kesamaan maksimum dan data antar *cluster* mempunyai suatu kemiripan minimum. partisi dalam clustering dilakukan secara otomatis dengan menggunakan algoritma *K-Means* (Irwansyah, 2015).

K-Means adalah algoritma yang masuk ke dalam kategori Unsupervised learning, Unsupervised learning digunakan sebagai pembagi data menjadi suatu kelompok data dengan menggunakan sytem partisi (Wanto, 2020). Berikut merupakan alur dalam melakukan clustering metode *K-Means* sebagai berikut:

1. Pilih (tentukan) nilai "*K*" (*cluster*) yang akan dipakai.
2. Menentukan nilai pusat *centroid* awal secara acak (*random*).
3. Hitung jarak pada setiap *input* data terhadap masing-masing *centroid* dengan menggunakan rumus *euclidean distance* sebagai berikut:

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2}$$

Keterangan:

$d(x, y)$ = ukuran ketidakmiripan

x_i = (x_1, x_2, \dots, x_i) yaitu variabel data

y_i = (y_1, y_2, \dots, y_i) merupakan variabel *centroid*

4. Klasifikasikan data menurut jarak kedekatannya dengan *centroid* (jarak terkecil).
5. Hitung kembali (memperbarui) nilai *centroid* pada iterasi baru dengan menggunakan rumus sebagai berikut:

$$c_{m(q)} = \frac{1}{n_m} \sum_{i=1}^{n_m} x_i(q)$$

Keterangan:

$C_{m(q)}$ = *centroid cluster*

n_m = jumlah data yang menjadi anggota *cluster* ke- i

m, i = indeks dari *cluster*

$x_i(q)$ = nilai data dari ke- i yang berada di dalam *cluster* tersebut untuk variabel ke- q

6. Lakukan perhitungan ulang pada langkah 3 hingga 5 sampai anggota pada tiap *cluster* tidak mengalami perubahan, maka proses *clustering* sudah selesai.
7. Apabila langkah 6 telah selesai maka nilai pusat *cluster* pada iterasi terakhir akan dipakai untuk parameter menentukan pengelompokan data.

Setelah proses *clustering* selesai maka langkah terakhir merupakan tahap evaluasi *cluster* yang bertujuan untuk mengetahui seberapa baik suatu data yang sudah dikelompokkan menggunakan metode *Davies-Bouldin Index* (DBI) dengan standar nilai indikator bahwa semakin kecil nilai dari DBI yang didapat (non-negatif ≥ 0), maka semakin baik *cluster* yang didapat dari pengelompokan *cluster* yang diterapkan.

HASIL DAN DISKUSI

Sebelum proses *clustering* ini dilakukan akan terlebih dahulu dilakukan tahap *preprocessing* data yang meliputi *data reduction*, *data cleaning*, *data transformation*, dan *data integration* sebagai berikut:

a. Data Reduction

Data reduction merupakan proses untuk melakukan pengurangan data apabila ada data yang kurang sesuai maupun data yang tidak digunakan dalam penelitian.

b. Data Cleaning

Dalam melakukan *data cleaning* digunakan sebagai mengisi value yang kosong dalam penulisan dan melakukan koreksi terhadap data yang redundansi

c. Data Transformation

Setelah melalui tahap *data reduction* dan *data cleaning* penulis sudah memiliki data yang sesuai dengan kebutuhan dan didapatkan data rekam medis pasien rawat jalan dan rawat inap pada bulan Agustus-Oktober tahun 2021 sebanyak 369 data seperti pada tabel 1 berikut ini:

Tabel 1. Sampel Data Rekam Medis

Data Pasien	Kecamatan	Diagnosa
PSN 1	Tulangan	A09
PSN 2	Tulangan	A09
PSN 3	Tulangan	A09
PSN 4	Tulangan	A09
PSN 5	Tulangan	A09
PSN 6	Tanggulangin	A09
PSN 7	Tulangan	A09
PSN 8	Krembung	A09
PSN 9	Tanggulangin	A09.9
PSN 10	Tulangan	A15.3
⋮	⋮	⋮
PSN 369	Tulangan	I11.0

Pada tabel 1. Peneliti menggunakan 2 variabel dalam rekam medis berupa kecamatan, dan diagnosa menurut terminologi ICD-10 [15]. Langkah selanjutnya dalam melakukan transformasi data adalah dengan cara melakukan inialisasi pada setiap data yang mempunyai *type* data *non numerik* menjadi data *numerik* seperti jenjang umur, jenis kelain, kecamatan, dan diagnosa yang bertujuan untuk mempermudah proses perhitungan dalam *K-Means clustering*. Berdasarkan data pada tabel 1, Berikut merupakan hasil inialisasi data untuk variabel kecamatan, dan diagnosa ICD-10 yang dapat dilihat pada tabel 2, tabel 3, tabel 4, dan tabel 5 sebagai berikut:

Tabel 2. Inialisasi Wilayah (Kecamatan)

Kecamatan	Inialisasi
Buduran	1
Candi	2
Gedangan	3
Jabon	4
Krembung	5
Krian	6
Prambon	7
Porong	8
Sedati	9
Sidoarjo	10
Sukodono	11
Taman	12
Tanggulangin	13
Tarik	14
Tulangan	15
Waru	16
Wonoayu	17
Balongbendo	18

Tabel 3. Inisialisasi Kode Diagnosa ICD-10

Kode ICD-10	Terminologi ICD-10	Inisialisasi
A09	<i>Other gastroenteritis and colitis of infectious and unspecified origin</i>	1
A09.0	<i>Gastroenteritis and colitis of unspecified origin</i>	2
A15.3	<i>Tuberculosis of lung, confirmed by unspecified means</i>	6
A20.2	<i>Pneumonic plague</i>	8
A91	<i>Dengue haemorrhagic fever</i>	11
B16	<i>Acute hepatitis B</i>	12
B19	<i>Unspecified viral hepatitis</i>	14
B36.9	<i>Histoplasmosis</i>	19
B90	<i>Sequelae of tuberculosis</i>	21
B90.9	<i>Sequelae of respiratory and unspecified tuberculosis</i>	22
C73	<i>Malignant neoplasm of thyroid gland</i>	29
D14.4	<i>Respiratory system, unspecified</i>	32
D18.0	<i>Haemangioma, any site</i>	34
D21	<i>Other benign neoplasms of connective and other soft tissue</i>	35
D21.9	<i>Connective and other soft tissue, unspecified</i>	38
D24	<i>Benign neoplasm of breast</i>	39
D64	<i>Other anaemias</i>	42
E03.9	<i>Hypothyroidism, unspecified</i>	45
E05.9	<i>Thyrotoxicosis, unspecified</i>	46
E06.9	<i>Thyroiditis, unspecified</i>	47
E10.8	<i>Type 1 diabetes mellitus With unspecified complications</i>	50
E11.6	<i>Type 2 diabetes mellitus With other specified complications</i>	55
E11.8	<i>Type 2 diabetes mellitus With unspecified complications</i>	56
E25.0	<i>Congenital adrenogenital disorders associated with enzyme deficiency</i>	59
E43	<i>Unspecified severe protein-energy malnutrition</i>	60
E46	<i>Unspecified protein-energy malnutrition</i>	61
F80.9	<i>Developmental disorder of speech and language, unspecified</i>	63
G35	<i>Multiple sclerosis</i>	64
G40	<i>Epilepsy</i>	65
G83.3	<i>Monoplegia, unspecified</i>	67
G96.0	<i>Cerebrospinal fluid leak</i>	68
H10.1	<i>Acute atopic conjunctivitis</i>	72
H60	<i>Otitis externa</i>	75
H61.2	<i>Impacted cerumen</i>	77
H65.0	<i>Acute serous otitis media</i>	78
H65.2	<i>Chronic serous otitis media</i>	79
H68.1	<i>Obstruction of Eustachian tube</i>	81
H81.1	<i>Benign paroxysmal vertigo</i>	83

Kode ICD-10	Terminologi ICD-10	Inisialisasi
H93.1	<i>Tinnitus</i>	84
I05.1	<i>Rheumatic mitral insufficiency</i>	85
I05.2	<i>Mitral stenosis with insufficiency</i>	86
I10	<i>Essential (primary) hypertension</i>	88
I10.0	<i>Hypertensive heart disease with (congestive) heart failure</i>	90

Setelah proses transformasi data selesai maka hasil dari inisialisasi dapat dilihat pada tabel 4 berikut ini:

Tabel 4. Hasil Transformasi Sampel Data Rekam Medis

Data Pasien	Kecamatan	Diagnosa
PSN 1	15	1
PSN 2	15	1
PSN 3	15	1
PSN 4	15	1
PSN 5	15	1
PSN 6	13	1
PSN 7	15	1
PSN 8	5	1
PSN 9	13	1
PSN 10	15	2
⋮	⋮	⋮
PSN 369	15	90

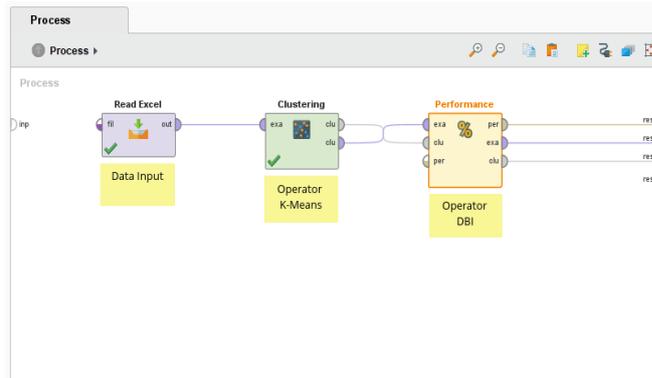
d. *Data Integration*

Data integration digunakan untuk menggabungkan 2 *file* yang berbeda. Pada penelitian ini data yang digunakan bersumber dari 2 data *file* kunjungan rawat jalan dan rawat inap yang berbeda dan disatukan menjadi satu tabel dalam *microsoft excel* agar mempermudah saat melakukan perhitungan menggunakan algoritma *K-Means Clustering*. Setelah melakukan tahap *preprocessing* dan mendapatkan *dataset* yang siap untuk diolah maka langkah berikutnya merupakan proses untuk menerapkan algoritma *K-Means* menggunakan *software* RapidMiner studio

e. *Implementasi Algoritma K-Means Menggunakan Software* RapidMiner

Perhitungan pengelompokan data rekam medis menggunakan RapidMiner bertujuan untuk mengekstrak pola-pola dalam *dataset* menggunakan rumus kombinasi dengan metode *artificial intelligence*, statistika, dan *database*.

Pada penelitian ini, *dataset* yang dipakai dalam *software* RapidMiner merupakan *dataset* yang berasal dari *microsoft excel*. Setelah itu pilih *file dataset* yang akan dipakai, *import file* yang sudah dipilih, kemudian tentukan indikator tipe *data* pada masing-masing *attribute* dengan ketentuan apabila pada setiap *attribute* berisikan data berupa numerik maka *type data* yang digunakan adalah *integer*, sedangkan apabila tipe *data* yang digunakan berisikan data berupa *alfanumerik* maka *type data* yang digunakan adalah *polynomial*. Berikut merupakan cara pengaturan *dataset* dalam melakukan *import* dan desain ke dalam *software* RapidMiner seperti pada gambar 2 berikut ini:



Gambar 1. Desain Lembar Kerja *Clustering* Rapidminer

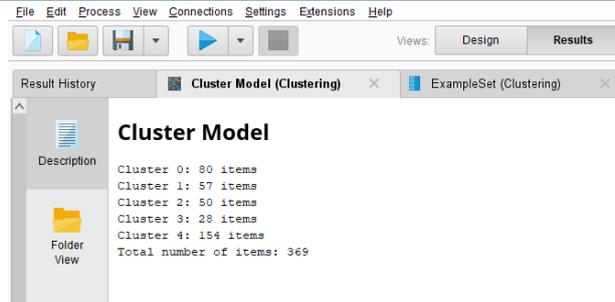
Pada gambar 1 diatas dapat dilihat bahwa setelah desain lembar kerja dibuat. Proses selanjutnya adalah pengaturan pada menu *parameters* operator *clustering*. *Parameters* yang diatur adalah nilai “K”, yang dimana “K” merupakan jumlah *cluster* yang akan dipakai. Dalam penelitian ini jumlah “K” yang dipakai adalah 5 *cluster* dan *max runs* berjumlah 10 yang berarti proses akan dilakukan *iterasi* sebanyak 10 kali. Sedangkan *numerical measure* menggunakan *euclidean distance* sebagai perhitungan jarak antar objek data dengan *centroid* yang digunakan.

Setelah proses pengaturan pada menu *parameters* selesai. Maka, proses selanjutnya merupakan tahap *running* data yang akan dilakukan dengan cara klik “F11” dan apabila proses *running* berhasil berjalan dengan lancar maka akan mendapatkan *output ExampleSet* berupa: *view data* seperti pada gambar 4 berikut ini:

Row No.	DATA PASIEN	cluster	UMUR	JENIS KELA...	KECAMATAN	DIAGNOSA
1	PSN 1	cluster_2	1	2	15	1
2	PSN 2	cluster_2	1	1	15	1
3	PSN 3	cluster_2	1	1	15	1
4	PSN 4	cluster_2	1	2	15	1
5	PSN 5	cluster_2	1	1	15	1
6	PSN 6	cluster_2	1	1	13	1
7	PSN 7	cluster_2	1	2	15	1
8	PSN 8	cluster_2	1	1	5	1
9	PSN 9	cluster_2	1	2	13	1
10	PSN 10	cluster_2	1	2	15	2
11	PSN 11	cluster_2	1	1	15	6
12	PSN 12	cluster_2	2	1	15	6
13	PSN 13	cluster_2	2	1	15	6
14	PSN 14	cluster_2	3	1	15	6
15	PSN 15	cluster_2	2	2	15	6

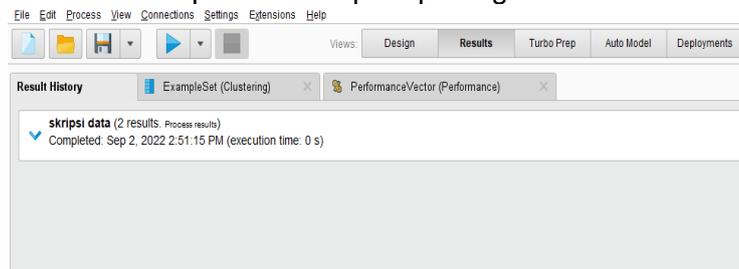
Gambar 2. Hasil *Clustering* Pada *Data View*

Setelah proses *running* berjalan dengan baik maka akan muncul tampilan pada *data view* seperti pada gambar 2 diatas dan didapatkan juga informasi hasil *cluster* yang digunakan dalam penelitian ini yaitu: C_0, C_1, C_2, C_3 , dan C_4 . Yang dimana C_0 mewakili *cluster* 0, C_1 mewakili *cluster* 1, C_2 mewakili *cluster* 2, C_3 mewakili *cluster* 3, dan C_4 mewakili *cluster* 4 dengan hasil seperti pada gambar 5 berikut ini:



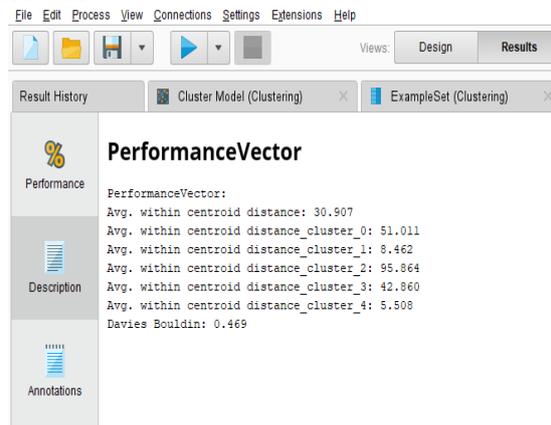
Gambar 3. Hasil *Cluster Model K-Means*

Pada gambar 3 diatas merupakan hasil dari proses perhitungan dari tiap anggota *cluster* dengan jumlah C0 adalah 80 items, C1 berjumlah 57 items, C2 berjumlah 50 items, C3 berjumlah 28 items, dan C4 berjumlah 154 items. Selanjutnya merupakan hasil lama waktu perhitungan algoritma *K-Means* menggunakan *software* RapidMiner seperti pada gambar 4 berikut ini:



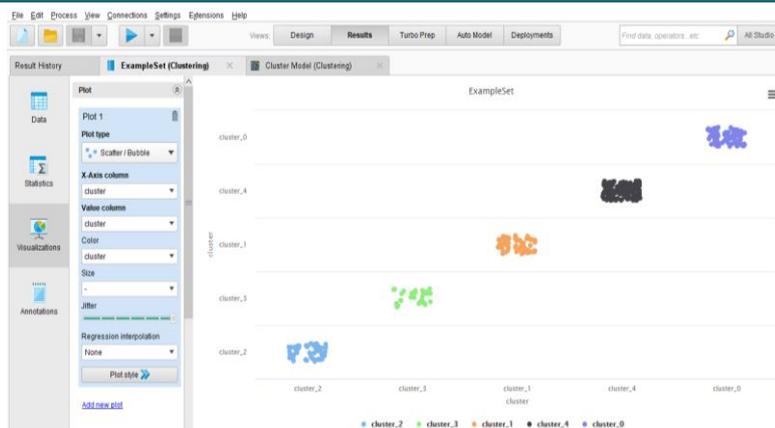
Gambar 4. *Result History* Menggunakan Algoritma *K-Means*

Pada gambar 4 di atas merupakan hasil dari waktu tunggu proses perhitungan yang dimulai dari awal proses *clustering* sampai dengan hasil akhir proses evaluasi *clustering*. Hasil dari nilai DBI dan *performance* yang menghasilkan *centroid distance* pada setiap *cluster* mendapatkan nilai 0,469 yang dapat dilihat pada gambar 4 berikut ini:



Gambar 5. *Centroid Distance K-Means*

Berikut merupakan hasil akhir (*output*) dari proses *clustering* menggunakan *software* RapidMiner berupa *group* mana yang mendapatkan tempat *clustering* pada level 0,1,2,3, dan 4. Seperti pada gambar 5 berikut ini:



Gambar 6. Hasil *Visualizations* Pada Tiap *Cluster*

Pada gambar 6 diatas merupakan *output* tampilan akhir dari perhitungan algoritma *K-Means* menggunakan *software* RapidMiner. Didapatkan hasil berupa, *cluster* 0 berjumlah 80 data yang dilambangkan dengan warna *dark blue*, *cluster* 1 berjumlah 57 data dilambangkan dengan warna *orange*, *cluster* 2 berjumlah 50 data dilambangkan dengan warna *light blue*, *cluster* 3 berjumlah 28 data dilambangkan dengan warna *green*, dan *cluster* 4 berjumlah 154 data dilambangkan dengan warna *black*.

UCAPAN TERIMA KASIH

Ucapan terima kasih ditujukan kepada Universitas Muhammadiyah Sidoarjo dan RS Aisyiyah Siti Fatimah Tulangan Sidoarjo atas ijinnya dalam kegiatan penelitian ini.

REFERENSI

- Adhi Kusumastuti, Ahmad Mustamil Khoiron, & Taofan Ali Achmadi. (2020). *Metode Penelitian Kuantitatif*. Deepublish.
- Amir Ali. (n.d.). *Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo*. Retrieved September 23, 2022, from <https://journal.universitاسbumigora.ac.id/index.php/matrik/article/view/529>
- Irwansyah, E. F. (2015). *Advanced Clustering Teori dan Aplikasi* (Yogyakarta). Penerbit Deepublish. //portal.upj.ac.id/slims/index.php?p=show_detail&id=9309&keywords=
- Kemendes RI. (2008). *Peraturan Menteri Kesehatan Republik Indonesia No. 269/MENKES/PER/III/2008*. <https://www.regulasip.id/book/4974/read>
- Kemendes RI. (2009). *UU No. 44 Tahun 2009 tentang Rumah Sakit*. <https://peraturan.bpk.go.id/Home/Details/38789/uu-no-44-tahun-2009>
- Mochamad Wahyudi,dkk. (n.d.). *Data mining penerapan algoritma k- means clustering dan k- medoids clustering—2020*. Retrieved September 23, 2022, from <https://elibrary.bsi.ac.id/readbook/212222/data-mining-penerapan-algoritma-k-means-clustering-dan-k-medoids-clustering>
- Mustika Mustika, Yunita Ardilla, Abraham Manuhutu, Nazaruddin Ahmad, Imanuddin Hasbi, Guntoro Guntoro, Melda Agnes Manuhutu, Mohamad Ridwan, Hozairi Hozairi, Anindya Khrisna Wardhani, Syariful Alim, Ikhsan Romli, Yoga Religia, D. Tri Octafian, Unggul Utan Sufandi, & Iin Ernawati. (2021). *DATA MINING DAN APLIKASINYA* (N. Rismawati, Ed.). Widina Bhakti Persada.
- Ordila, R., Wahyuni, R., Irawan, Y., & Yulia Sari, M. (2020). PENERAPAN DATA MINING UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT DENGAN ALGORITMA CLUSTERING (Studi Kasus: Poli Klinik

- PT.Inecda). *Jurnal Ilmu Komputer*, 9(2), 148–153.
<https://doi.org/10.33060/JIK/2020/Vol9.Iss2.181>
- Wanto, A. (2020). *Data Mining: Algoritma dan Implementasi*. Yayasan Kita Menulis.
- Wijaya, A. P., & Vera, O. (2021). Analisa dan Perencanaan Sistem Informasi Rekam Medis Pada Fasilitas Kesehatan. *Jurnal Ilmiah Infokam*, 17(2), 121–128.
<https://doi.org/10.53845/infokam.v17i2.305>