
CLUSTERING PENYAKIT DBD PADA RUMAH SAKIT DHARMA KERTI MENGGUNAKAN ALGORITMA K-MEANS

I Nyoman Mahayasa Adiputra^{1,*}

¹ Universitas Bali Internasional, Fakultas Ilmu – Ilmu Kesehatan, Program Studi Manajemen Informasi Kesehatan,, Jln. Seroja Gg Jeruk No. 9 Denpasar 80239 INDONESIA

Abstrak

Clustering merupakan teknik pengelompokan *record* pada basis data berdasarkan kriteria tertentu. Hasil *clustering* diberikan kepada pengguna akhir untuk memberikan gambaran tentang apa yang terjadi pada basis data. Dalam SIMRS Rumah Sakit Dharma Kerti yang diambil dari tahun 2014 sampai 2018, memiliki 6038 *record* data rekam medis pasien. Jumlah diagnosa tertinggi adalah DBD, untuk mengetahui gambaran yang terjadi dalam data diagnosa DBD tersebut dilakukan *clustering* dengan menggunakan algoritma *k-means*. Hasil dari penelitian ini adalah pasien dengan diagnosa DBD pada RS Dharma Kerti yang paling sedikit jumlahnya adalah pada rentang usia 25 sampai 45 tahun. Setiap kelompok usia memiliki jumlah hari perawatan yang merata, bisa disimpulkan bahwa usia hampir tidak mempengaruhi hari perawatan. Lama perawatan yang paling banyak dilalui oleh pasien adalah dari 3 hari sampai 5 hari perawatan.

Kata Kunci:

Clustering, Diagnosa DBD, *k-means*

Abstract

Clustering is a technique for grouping records in the database based on certain criteria. The results of the clustering are given to the end user to give an idea of what is happening to the database. In Dharma Kerti Hospital Information System which was taken from 2014 to 2018, it has 6038 records of patient medical record data. The highest number of diagnoses is DHF, to find out the picture that occurs in the DHF diagnostic data, clustering is done using the *k-means* algorithm. The results of this study are patients with a diagnosis of dengue fever at Dharma Kerti Hospital, the least in number are in the age range of 25 to 45 years. Each age group has an even number of treatment days, it can be concluded that age does not affect treatment days. The length of treatment that most patients go through is from 3 days to 5 days of treatment

Keywords:

clustering, DHF Diagnostic, *K-means*

1. PENDAHULUAN

Clustering merupakan teknik pengelompokan *record* pada basis data berdasarkan kriteria tertentu. Hasil *clustering* diberikan kepada pengguna akhir untuk memberikan gambaran tentang apa yang terjadi pada basis data. Konsep dasar dari *clustering* adalah mengelompokkan sejumlah objek ke dalam *cluster* dimana *cluster* yang baik adalah *cluster* yang memiliki tingkat kesamaan yang tinggi antar objek di dalam suatu *cluster* dan tingkat ketidaksamaan yang tinggi dengan objek *cluster* yang lainnya. Metode *clustering* hadir dikarenakan kondisi berlimpahnya data yang merupakan akumulasi data transaksi yang terekam selama bertahun-tahun (Ali, 2020).

Dalam SIMRS Rumah Sakit Dharma Kerti yang diambil dari tahun 2014 sampai 2018, memiliki 6038 *record* data rekam medis pasien. Dari hasil penelitian sebelumnya jumlah data penyakit tertinggi untuk seluruh tahun adalah DBD atau Demam Berdarah Dengue dengan total *record* 1323 data rekam medis pasien (Adiputra & Darmayanti, 2020). Dengan jumlah data penyakit tertinggi adalah DBD maka dalam penelitian ini akan mengimplementasikan algoritma *cluster* yaitu *K-Means clustering* untuk mengelompokkan data rekam medis pasien dalam penyakit DBD. Teknik *clustering* dibutuhkan berdasarkan data dari penelitian sebelumnya (Adiputra & Darmayanti, 2020), penyakit DBD merupakan salah satu

penyakit tertinggi pada RS Dharma Kerti yang terjadi pada setiap tahunnya, dengan melakukan *clustering* pada data penyakit DBD RS dapat mengetahui informasi yang mendalam mengenai penyakit DBD yang terjadi, sehingga RS dapat mengambil keputusan yang tepat berdasarkan data.

K-Means Clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu cluster/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan cluster/kelompok yang lain sehingga data yang berada dalam satu cluster/kelompok memiliki tingkat variasi yang kecil. K-Means merupakan suatu algoritma yang digunakan dalam pengelompokkan secara partisi yang memisahkan data ke dalam kelompok yang berbeda – berbeda. Algoritma ini mampu meminimalkan jarak antara data ke clusternya. Pada dasarnya penggunaan algoritma ini dalam proses clustering tergantung pada data yang didapatkan dan konklusi yang ingin dicapai di akhir proses (Gustientiedina, Adiya, & Desnelita, 2019).

Berdasarkan penelitian sebelumnya yang berjudul “Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Cigugur Tengah” pada tahun 2020, Puskesmas Cigugur Tengah dalam setiap harinya melayani pasien sekitar 150 orang dari berbagai wilayah di 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 mengelompokkan 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 pembanding. 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 (Sugianto, Rahayu, & Gusman, 2020).

Terdapat pula penelitian dengan judul “Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka)” yang menyatakan Algoritma K-means tidak terpengaruh terhadap urutan objek yang digunakan, hal ini dibuktikan ketika penulis mencoba menentukan secara acak titik awal pusat cluster dari salah satu objek pada permulaan perhitungan. Jumlah keanggotaan cluster yang dihasilkan berjumlah sama ketika menggunakan objek yang lain sebagai titik awal pusat cluster tersebut. Namun, hal ini hanya berpengaruh pada jumlah iterasi yang dilakukan. Pengelompokan objek (objek clustering) adalah salah satu proses dari objek mining yang bertujuan untuk mempartisi objek yang ada ke dalam satu atau lebih cluster objek berdasarkan karakteristiknya. Penelitian ini mengkaji bagaimana penggunaan Algoritma K-means Cluster Analysis dalam studi kasus penyakit menular manusia pada suatu objek. Penelitian ini mengkaji metode K-means Cluster Analysis dalam penyakit menular pada manusia berdasarkan set variabel yang dibentuk per kecamatan tiap Puskesmas yang jumlahnya ada 32 Kantor Puskesmas di Kabupaten Majalengka (Bastian, Sujadi, & Febrianto, 2018).

Penelitian lain dengan judul “Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau” pada tahun April 2021 mendapatkan hasil algoritma K-Means mampu mengelompokkan penyebab penyebaran penyakit ISPA ke dalam 2 cluster, yang mana cluster 1 memberikan rekomendasi tinggi berjumlah 10 Kabupaten, cluster 2 memberikan rekomendasi rendah berjumlah 2 Kabupaten. Dalam pengujian ini dilakukan dengan menggunakan dua aplikasi, yang pertama dengan menggunakan aplikasi pengolah angka Microsoft Excel dan dengan menggunakan aplikasi Rapidminer 5.3 menggunakan K-Means dapat menampilkan 2 cluster dari hasil klasifikasi (Bastian et al., 2018).

Berdasarkan hasil penelitian – penelitian sebelumnya algoritma k-means memiliki hasil yang lebih baik dibandingkan algoritma lain, dan juga algoritma k-means mampu melakukan pengelompokan terhadap data penyakit, penelitian ini akan menggunakan algoritma K-Means untuk mengetahui cluster dari data rekam medis pada penyakit DBD dalam Rumah Sakit Dharma Kerti dari tahun 2014 sampai 2018, hasil pengelompokan atau cluster data tersebut nantinya dapat dijadikan analisis untuk mengetahui pasien dengan variabel (usia, jenis kelamin, periode masuk RS, Kelas Kamar) seperti apa yang datang ke Rumah Sakit Dharma Kerti. Data hasil cluster dapat dimanfaatkan oleh Rumah Sakit dalam teknologi berbasis

website sehingga dapat diakses secara fleksibel sebagai decision support atau membantu dalam membuat kebijakan (Teufel & Binder, 2021) terhadap fenomena penyakit DBD dalam Rumah Sakit Dharma Kerti.

2. METODE

A. Data Mining

Data mining dapat diartikan sebagai proses mengekstrak atau menggali knowledge yang ada pada sekumpulan data. Informasi dan knowledge yang didapat tersebut dapat digunakan pada banyak bidang, seperti manajemen bisnis, pendidikan, kesehatan dan sebagainya. data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari database yang besar. Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Proses menggali informasi dalam data mining melibatkan integrasi teknik dari berbagai disiplin ilmu, seperti teknologi database dan *data warehouse*, statistik, *machine learning*, komputasi dengan kinerja tinggi, pattern recognition, neural network, visualisasi data dan sebagainya (Ediyanto, Mara, & Satyahadewi, 2013).

B. Clustering

Metode Clustering adalah mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan kedalam satu cluster yang sama. Tujuan dari clustering ini adalah untuk meminimalisir fungsi tujuan yang diterapkan dalam proses clustering. Yang umumnya berusaha meminimalisir variasi dalam suatu cluster dan memaksimalkan variasi antar cluster (Purba, Poningsih, & Tambunan, 2021). Dalam penelitian ini metode utama yang digunakan dalam pengolahan data adalah *clustering*, data rekam medis yang didapatkan dan diolah dalam Bahasa pemrograman menggunakan Teknik *clustering*

C. Algoritma *K-means*

Algoritma *K-Means* merupakan algoritma klusterisasi yang mengelompokkan data berdasarkan titik pusat klaster (*centroid*) terdekat dengan data. Tujuan *K-Means* adalah pengelompokkan data dengan memaksimalkan kemiripan data dalam satu klaster dan meminimalkan kemiripan data antara klaster. ukuran kemiripan yang digunakan dalam klaster adalah fungsi jarak. sehingga pemaksimalan kemiripan data didapatkan berdasarkan jarak terpendek antara data titik *centroid*.

Tahapan algoritma k-means adalah sebagai berikut :

- a) Pilih secara acak k buah data sebagai pusat cluster.
- b) Jarak antara data dan pusat cluster dihitung menggunakan Euclidian Distance. Untuk menghitung jarak semua data ke setiap titik pusat cluster dapat menggunakan teori jarak Euclidean yang dirumuskan sebagai berikut:

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

dimana:

d = Jarak data ke pusat cluster
 x1= titik pertama pada atribut x
 x2 = titik kedua pada atribut x
 y1= titik pertama pada atribut y
 y2 = titik kedua pada atribut y

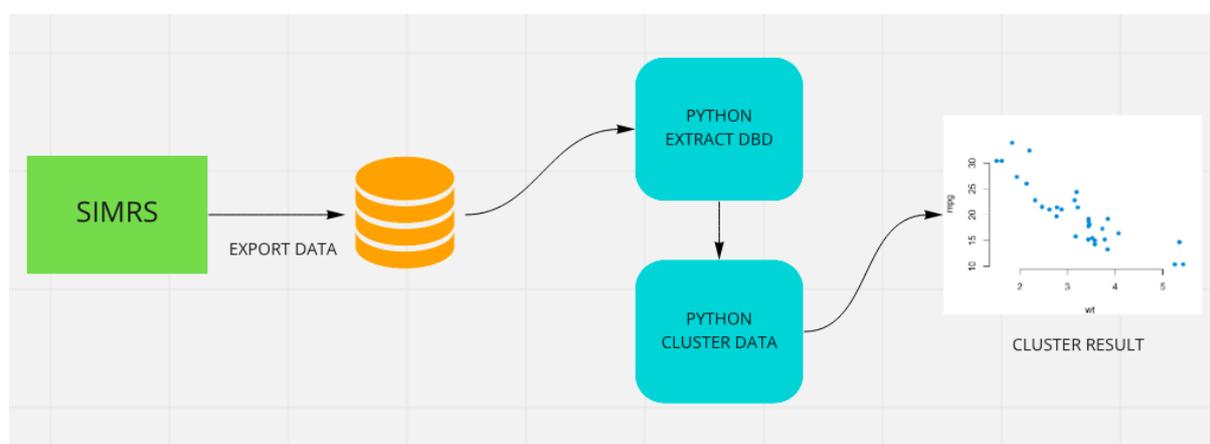
- c) Data ditempatkan dalam cluster yang terdekat, dihitung dari tengah cluster.
- d) Pusat cluster baru akan ditentukan bila semua data telah ditetapkan dalam cluster terdekat.
- e) Proses penentuan pusat cluster dan penempatan data dalam cluster diulangi sampai nilai centroid tidak berubah lagi

D. Pengumpulan Data

Penelitian ini merupakan penelitian deskriptif kuantitatif melalui pendekatan cross sectional yaitu variabel penelitian diukur atau dikumpulkan dalam satu waktu artinya mengadakan pengamatan hanya sekali terhadap beberapa variabel dan waktu bersamaan. Ruang lingkup penelitian ini adalah data digital pada SIMRS rekam medis dengan diagnosa DBD.

Penelitian ini akan melakukan implementasi terhadap algoritma *K-Means clustering* terhadap data rekam medis pasien dengan diagnosa DBD. Populasi target penelitian ini adalah rekam medis di rumah sakit sedangkan populasi terjangkanya adalah rekam medis digital pasien di Rumah Sakit Dharma Kerti. Sampel pada penelitian ini adalah data rekam medis digital pasien dengan diagnosa DBD di Rumah Sakit Dharma Kerti Tabanan dari tahun 2014 - 2018. Analisis yang dilakukan dalam penelitian ini dilakukan secara deskriptif terhadap hasil dari *K-Means clustering* pada data rekam medis dengan diagnosa DBD pada di Rumah Sakit Dharma Kerti.

E. Tahapan Cluster



Gambar 1. Tahapan *cluster*

Gambar 1 merupakan tahapan *cluster* yang dipakai pada penelitian ini, sumber data di dapat dari hasil export data pada SIMRS yang digunakan oleh RS Dharma kerti. Setelah mendapatkan data yang berbentuk SQL, data di konversi menjadi CSV kemudian data dalam bentuk CSV di import ke dalam Bahasa pemrograman python. Tahap pertama dalam pemrograman adalah melakukan ekstraksi penyakit DBD dengan menggunakan kode ICD 10 : A91, setelah mendapatkan data penyakit DBD tahap berikutnya adalah clustering masih dengan Bahasa pemrograman python. Terdapat dua parameter yang digunakan dalam clustering yaitu umur pasien dan berapa hari perawatan yang di dapat oleh pasien dari mulai masuk RS sampai keluar RS, tujuan dari kedua parameter tersebut untuk mengetahui apakah usia pasien DBD pada RS Dharma kerti mempengaruhi lama rawatnya dalam RS.

3. HASIL DAN PEMBAHASAN

Penelitian ini memanfaatkan teknologi Bahasa pemrograman python untuk memperoleh hasil cluster dari penyakit DBD. Langkah pertama yang dilakukan adalah input data dengan field data yaitu : *gender* (jenis kelamin), umur, diagnosa, lama rawat seperti yang terlihat pada Gambar 2 dibawah ini.

```
In [2]: df = pd.read_csv('../input/dhf-diagnosis-data/fact_simrm.csv')
df.head()
```

```
Out[2]:
```

	gender	umur	diagnosa	lama
0	M	37	A91	4
1	M	18	A91	4
2	M	30	A91	4
3	F	37	A91	6
4	F	16	A91	6

Gambar 2. Input data

Dari hasil input data dilakukan *cluster* berdasarkan umur pasien dan berapa lama perawatan pasien dengan diagnosa DBD di Rumah Sakit, jumlah *cluster* yang dilakukan pada penelitian ini adalah 3 *cluster*. Penggunaan 3 *cluster* didasarkan pada perhitungan SSE atau *Sum of Squared Error* yang dapat dilihat pada Gambar 5.

```
In [4]: km = KMeans(n_clusters=3)
y_predicted = km.fit_predict(df[['umur', 'lama']])
y_predicted
```

```
Out[4]: array([2, 0, 2, ..., 2, 0, 0], dtype=int32)
```

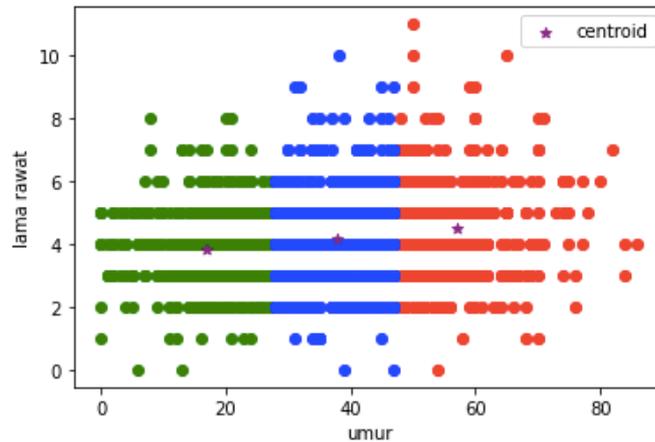
```
In [5]: df['cluster']=y_predicted
df.head()
```

```
Out[5]:
```

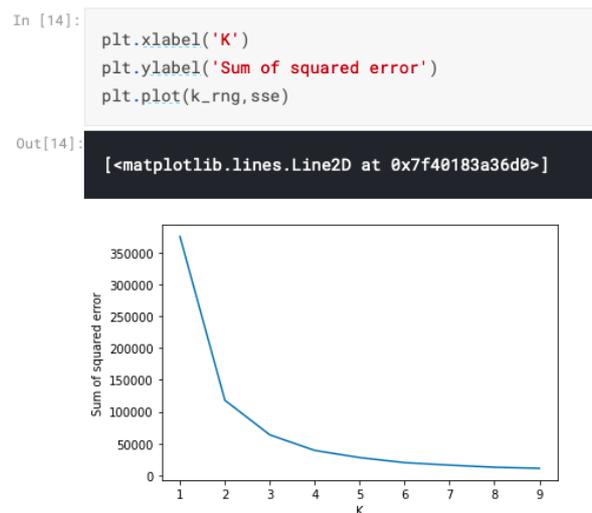
	gender	umur	diagnosa	lama	cluster
0	M	37	A91	4	2
1	M	18	A91	4	0
2	M	30	A91	4	2
3	F	37	A91	6	2
4	F	16	A91	6	0

Gambar 3. Proses *clustering*

Hasil *cluster* dalam bentuk tabel dapat dilihat pada Gambar 3, dalam tabel tersebut di tambahkan satu *field* untuk menampung hasil *cluster*, dari tabel yang dihasilkan tersebut langkah berikutnya adalah menampilkan visualisasi hasil *cluster*, untuk memudahkan dalam membaca data dan memberikan analisis dari hasil *cluster*.

Gambar 4. Hasil *clustering*

Visualisasi hasil *cluster*, dapat dilihat pada Gambar 4 diatas. Sesuai dengan input yang di berikan pada inisialisasi awal *cluster* akan terbagi menjadi 3 sesuai dengan warna yang terlihat yaitu hijau, biru dan merah dengan tanda bintang sebagai *centroid* atau titik tengah dari *cluster*



Gambar 5. Sum of squared error

K-means memiliki kelemahan yang diakibatkan oleh penentuan pusat awal cluster. Hasil cluster yang terbentuk dari metode K-means ini sangatlah tergantung pada inisiasi nilai pusat awal cluster yang diberikan. Untuk menghasilkan pusat jumlah cluster yang optimal digunakan metode SSE (Thinsungnoen, Kaoungku, Durongdumronchai, Kerdprasop, & Kerdprasop, 2015). Seperti pada Gambar 5 SSE yang optimal adalah pada angka 3 karena grafik sudah mulai menurun drastis pada angka 4, maka dalam penelitian ini jumlah cluster yang digunakan adalah 3.

Hasil dari penelitian ini dapat dilihat pada visualisasi cluster yang ada pada Gambar 3. Dari hasil visualisasi tersebut data diagnosa DBD pada Rumah Sakit untuk keseluruhan tahun dapat di bagi menjadi 3 kelompok berdasarkan usia atau umur. Kelompok pertama berwarna hijau adalah dari rentang usia 0 – 25, rentang usia kedua dari 26 sampai 45 dengan warna biru dan rentang usia ketiga dari 46 sampai diatas 80 dengan warna merah. Lama perawatan yang di dapatkan dari hasil cluster adalah di setiap rentang usia memiliki jumlah data yang hampir sama dalam lama perawatan, di mulai dari 2 hari sampai paling lama

adalah 10 hari. Jumlah lama perawatan yang paling banyak ada pada 3 hari sampai 5 hari, sedangkan di 8 hari ke atas data yang di dapatkan semakin berkurang.

4. SIMPULAN DAN SARAN

Kesimpulan yang di dapat dari hasil penelitian ini adalah Rumah Sakit memiliki pasien dengan diagnosa DBD yang terbagi menjadi 3 kelompok usia diantaranya adalah pasien dengan kategori usia 25 sampai 45 dengan warna biru memiliki jumlah diagnosa paling sedikit. Setiap kelompok usia memiliki jumlah hari perawatan yang merata, bisa disimpulkan bahwa usia hampir tidak mempengaruhi hari perawatan. Lama perawatan yang paling banyak di lalui oleh pasien adalah dari 3 hari sampai 5 hari perawatan. Penelitian kedepan dapat menambahkan variabel - variabel lain untuk mempelajari lebih dalam bagaimana pengelompokan penyakit DBD terhadap orang yang terjangkit, perbandingan algoritma selain k-means juga dapat menjadi konsen penelitian kedepan untuk mengetahui hasil *clustering* dari sisi algoritma.

Ucapan Terima Kasih

Ucapan terima kasih yang sebesar - besarnya peneliti sampaikan pada Rumah Sakit Dharma Kerti Tabanan - Bali yang telah memberikan kesempatan untuk melakukan penelitian dan memberikan data diagnosa sehingga penelitian ini dapat memberikan hasil yang bermanfaat untuk berbagai kalangan.

Daftar Pustaka

- Adiputra, I. N. M., & Darmayanti, N. L. (2020). Implementasi datawarehouse pada sistem informasi rumah sakit. *Jurnal Teknologi Informasi Dan Komputer*, 6(1), 78–85.
- Ali, A. (2020). Clustering Data Antropometri Balita Untuk Menentukan Status Gizi Balita Di Kelurahan Jumput Rejo Sukodono Sidoarjo. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(3), 395–407. <https://doi.org/10.35957/jatisi.v7i3.530>
- Bastian, A., Sujadi, H., & Febrianto, G. (2018). Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). *Jurnal Sistem Informasi (Journal of Information System)*, 14(1), 26–32.
- Ediyanto, Mara, N., & Satyahadewi, N. (2013). Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, 02(2), 133–136.
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Purba, N., Poningsih, P., & Tambunan, H. S. (2021). Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau. *Journal of Information System Research (JOSH)*, 2(3), 220–226. Retrieved from <http://ejurnal.seminar-id.com/index.php/josh/article/view/736>
- Sugianto, C. A., Rahayu, A. H., & Gusman, A. (2020). Algoritma K-Means untuk Pengelompokkan Penyakit Pasien pada Puskesmas Cigugur Tengah. *Journal of Information Technology*, 2(2), 39–44. <https://doi.org/10.47292/joint.v2i2.30>
- Teufel, A., & Binder, H. (2021). Clinical Decision Support Systems. *Visceral Medicine*, 37(6).
- Thinsungnoen, T., Kaoungku, N., Durongdumronchai, P., Kerdprasop, K., & Kerdprasop, N. (2015). The Clustering Validity with Silhouette and Sum of Squared Errors. In *3rd International Conference on Industrial Application Engineering* (pp. 44–51). <https://doi.org/10.12792/iciae2015.012>