

# IMPLEMENTASI K-MEANS DALAM PENGELOMPOKAN PENYEBARAN PENYAKIT DBD DI JAWA BARAT

Alpin Apriliansyah Mohsa<sup>1</sup>, Puput Silva Rosiana, Yuyun Umaidah<sup>3</sup>

<sup>1,2,3</sup> Universitas Singaperbangsa Karawang; Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Telp. (0267) 641177

---

## Riwayat artikel:

Received: 18 Juli 2023

Accepted: 30 Juli 2023

Published: 1 Agustus 2023

## Keywords:

3-5 keyword;

Algorithm a;

B algorithms;

Complexity.

## Correspondent Email:

[alpinmohsa@gmail.com](mailto:alpinmohsa@gmail.com)

© 2023 JITET (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY NC)

**Abstrak.** Penyakit menular ini disebut penyakit demam berdarah atau DBD adalah penyakit infeksi virus aedes aegypti yang akut yang disebabkan oleh virus dengue yang di tularkan melalui gigitan pada nyamuk. Berdasarkan tahun 2020, terjadi 24 ribu kasus infeksi DBD di Provinsi Jawa Barat. Maka dari itu, pada penelitian ini mengenai penyakit atau infeksi virus DBD pada Provinsi Jawa Barat menjadi sangat penting untuk dilakukan penelitian untuk menemukan pola penyebaran pada kasus ini. Data yang digunakan dalam penelitian ini berasal dari Open Data Jawa Barat, yaitu portal resmi data terbuka milik Pemerintah Provinsi Jawa Barat. Data yang diambil adalah data yang terkait dengan kasus infeksi Demam Berdarah Dengue (DBD) berdasarkan provinsi dari tahun 2014 hingga 2021. Open Data Jawa Barat menyediakan data dari Perangkat Daerah di lingkungan Pemerintah Provinsi Jawa Barat guna memenuhi kebutuhan masyarakat akan data yang akurat dan terkini. Kemudian data ini akan diseleksi, dari total terdapat 9 atribut hanya 3 atribut saja yang dipilih yaitu nama kabupaten kota, jumlah kasus, dan tahun. Jumlah anggota cluster 0 menunjukkan 139 data dan kriteria rendah 139 data, lalu untuk cluster 1 menunjukkan 33 data dengan kriteria tinggi 33 data, dan untuk cluster 2 memiliki 260 data dengan kriteria sedang 260 data. Evaluasi yang digunakan yaitu dengan teknik silhouette score dengan hasil score 0,5931974349414901.

**Abstract.** This infectious disease called dengue fever or DHF is an acute aedes aegypti virus infection caused by the dengue virus that is transmitted through mosquito bites. Based on 2020, there were 24 thousand cases of dengue infection in West Java Province. Therefore, in this study regarding the disease or dengue virus infection in West Java Province, it is very important to conduct research to find the distribution pattern in this case. The data used in this study comes from Open Data West Java, the official open data portal of the West Java Provincial Government. The data taken is data related to dengue fever infection cases by province from 2014 to 2021. Open Data West Java provides data from Regional Apparatus within the West Java Provincial Government to meet the public's need for accurate and up-to-date data. Then this data will be selected, from a total of 9 attributes only 3 attributes are selected, namely the name of the city district, the number of cases, and the year. The number of cluster 0 members shows 139 data and low criteria 139 data, then for cluster 1 shows 33 data with high criteria 33 data, and for cluster 2 has 260 data with medium criteria 260 data. The evaluation used is the silhouette score technique with a score of 0.5931974349414901.

## 1. PENDAHULUAN

Penyakit menular ini disebut penyakit demam berdarah atau DBD adalah penyakit infeksi virus aedes aegypti yang akut yang disebabkan oleh virus dengue yang ditularkan melalui gigitan pada nyamuk. DBD sering ditemukan di daerah tropis dan subtropis di seluruh dunia, termasuk di Jawa Barat.[1]. Pada tahun 2020, terjadi sekitar 24 ribu kasus infeksi virus dengue aedes aegypti di Jawa Barat.

Provinsi Jawa Barat diperkirakan akan mengalami puncak musim penghujan hingga April 2022. Selama musim ini, potensi penyebaran virus Demam Berdarah Dengue (DBD) meningkat secara signifikan [2]. Oleh karena itu, penelitian mengenai infeksi virus aedes aegypti ini pada provinsi Jawa Barat menjadi sangat penting untuk dilakukan penelitian untuk menemukan pola penyebaran. Pada penelitian sebelumnya yang telah dilakukan pada Provinsi Jawa Barat mengenai bagaimana distribusi dan kepadatan nyamuk vektor DBD di Kabupaten Ciamis berbeda pada daerah dengan ketinggian yang berbeda. Penelitian mengenai infeksi virus aedes aegypti ini pada Provinsi Jawa Barat dapat memberikan informasi penting mengenai penyebaran penyakit ini dan membantu dalam upaya pencegahan dan pengendaliannya.

## 2. TINJAUAN PUSTAKA

### 2.1 Data Mining

Untuk menemukan pola yang baru disebut metode data mining, metode ini adalah suatu metode untuk menemukan pola baru maupun tren, dan korelasi yang berharga dalam jumlah besar, Pada data mining ini metode yang dipakai yaitu teknologi pengenalan pola seperti teknik matematika dan statistik [3]. Data mining atau *Knowledge Discovery in Databases* (KDD) atau *pattern recognition* adalah proses mengekstrak informasi dari kumpulan data yang ada dengan berbagai teknik analisis. Data mining merupakan bagian dari disiplin ilmu *Data Science*. Terdapat beberapa istilah lain yang memiliki makna sama dengan data mining, yaitu ekstraksi pengetahuan, analisa data/pola, kecerdasan bisnis, data *archaeology*, dan data *dredging*. Teknik data mining digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Beberapa teknik yang digunakan dalam proses penambangan

data antara lain *predictive modeling*, *database segmentation*, *link analysis*, dan *text mining*

### 2.2 Algoritme K-Means Clustering

Algoritma ini termasuk clustering non hirarki selain itu akan melakukan pembagian data ke dalam satu maupun yang lebih dari pada klaster, selain itu k-means ini menjadikan data dengan karakteristik yang mirip akan dikelompokkan pada satu cluster atau pada kelompok dengan data yang sama[4].

Algoritme k-means adalah algoritme teknik clustering (pengklasteran) termasuk kelompok pemodelan unsupervised learning pada data mining dan digunakan dalam pengelompokan data dengan sistem partisi [5]. Metode k-means menjadi metode clustering yang paling umum dan sederhana, karena k-means memiliki kemampuan dalam pengelompokan data dengan jumlah besar dalam waktu sangat cepat dan efisien [6].

Metode k-means juga dikenal sebagai analisis kelompok, di mana setiap kelompok memiliki karakteristik yang mirip satu sama lain dan berbeda dengan kelompok lainnya. Tujuan dari algoritme k-means adalah meminimalkan perbedaan pada data dalam setiap kelompok dan memaksimalkan perbedaan data dengan kelompok yang lain.[7]

### 2.3 Sumber Dataset

Pada penelitian ini menggunakan data pada Open Data Jawa Barat, yaitu portal resmi data terbuka yang dimiliki oleh Pemprov Jawa Barat. Portal ini menyediakan dataset dari seluruh organisasi ataupun pemerintah daerah di Jawa Barat guna memenuhi kebutuhan masyarakat akan data yang akurat dan terkini. Pada penelitian ini data yang digunakan adalah data tentang kasus infeksi virus aedes agypti atau biasa di sebut (DBD) yang terjadi di provinsi Jawa Barat dalam rentang waktu tahun 2014 hingga 2021.

### 2.4 Metode Elbow

Metode *elbow* adalah teknik populer yang digunakan dalam analisis data dan pembelajaran mesin untuk menentukan jumlah klaster optimal dalam set data. Metode ini terutama dapat diterapkan dalam algoritme pengelompokan, seperti pengelompokan *k-means*. Teknik ini memanfaatkan jumlah rata-rata hasil perbandingan antara jumlah cluster

yang akan membentuk siku pada suatu titik. Dengan menggunakan metode ini, kita dapat mengetahui jumlah cluster terbaik pada suatu data dengan melihat perbandingan jumlah cluster yang akan membentuk siku pada suatu titik.[8]. Metode Elbow merupakan teknik dalam data mining yang digunakan untuk menentukan jumlah cluster terbaik dalam suatu data. Dalam metode ini, peneliti menentukan jumlah cluster terbaik dengan menambah nilai cluster pada model data. Terdapat beberapa metode yang dapat digunakan untuk menentukan jumlah cluster terbaik, seperti metode K-Means [9].

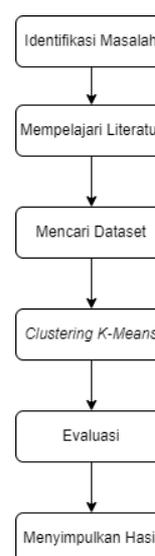
### 2.5 Evaluasi *Silhouette Coefficient*

Evaluasi *Silhouette Coefficient* adalah suatu metode dalam data mining yang digunakan untuk mengevaluasi kualitas cluster yang dihasilkan oleh algoritma clustering, seperti K-Means Clustering. Metode ini mengukur seberapa baik suatu objek ditempatkan dalam suatu cluster dengan memperhitungkan jarak antara objek tersebut dengan objek lain dalam cluster yang sama dan jarak antara objek tersebut dengan objek dalam cluster yang berbeda. Semakin tinggi nilai *Silhouette Coefficient*, semakin baik kualitas cluster yang dihasilkan[10].

### 3. METODE PENELITIAN

Penelitian ini menggunakan beberapa metode penelitian diantaranya pada gambar 1. Pada penelitian ini penulis menyajikan penelitian dengan eksperimen yang mampu mendapatkan permasalahan yang dihadapi melalui percobaan.

Tahapan-Tahapan pada penelitian ini diawali dengan identifikasi masalah lalu mempelajari literatur yang sudah dikumpulkan, mencari dataset untuk diolah, selanjutnya implementasi *Clustering k-means*, lalu setelah melewati tahap komputerisasi maka akan dilakukan evaluasi untuk mengetahui tingkat akurasi data yang diolah, dan yang terakhir menyajikan hasil penelitian.



Gambar 1. Alur Metode

## 4. HASIL DAN PEMBAHASAN

### 4.1 *Business Understanding*

Pada tahap *Business Understanding* ini yang dilakukan penulis untuk analisis yaitu dengan melakukan pemahaman bisnis serta penjelasan terhadap masalah data mining sehingga penulis dapat mudah untuk membuat penelitian dan dapat mencapai tujuan dan mendapatkan pola baru[11]. pada Pemahaman masalah penelitian ini mengacu pada penyebaran penyakit atau penyebaran infeksi virus aedes aegypti pada daerah Provinsi Jawa Barat. Periode musim hujan pada provinsi Jawa Barat pada puncaknya yaitu pada bulan April, maka dari itu pada musim penghujan penyakit atau virus DBD memiliki potensi penyebaran yang sangat tinggi. Untuk itu, untuk mengetahui titik rawan tingginya penyebaran penyakit DBD pada provinsi Jawa Barat, maka perlu dilakukannya pengelompokan daerah rawan penyebaran penyakit DBD pada provinsi Jawa Barat sebagai langkah awal meminimalisir bertambahnya masyarakat yang terkena penyakit atau virus DBD, dengan meningkatnya kasus pada setiap kasus pemerintah seharusnya sudah memiliki tindakan yang seharusnya untuk mengurangi terkenanya penyakit atau virus DBD ini pada provinsi Jawa Barat..

## 4.2 Data Understanding

Pada tahap data *understanding* ini menggunakan data penyebaran penyakit DBD pada provinsi Jawa Barat pada tahun 2014 - 2021 dan data ini disimpan dalam bentuk Microsoft Excel untuk memudahkan analisis dan pembacaan data pada penelitian. Jumlah data pada dataset penyebaran penyakit DBD pada tahun 2014 sampai dengan tahun 2021 ini dengan Jumlah data penyakit DBD di Provinsi Jawa Barat yaitu 432 data.

Pada penelitian ini variabel digunakan yaitu seperti pada Gambar 2 berikut ini contoh beberapa data pada dataset :

id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_kelamin	jumlah_kasus	satuan	tahun
1	32	JAWA BARAT	3201	KABUPATEN BOGOR	LAKI-LAKI	915	ORANG	2014
2	32	JAWA BARAT	3201	KABUPATEN BOGOR	PEREMPUAN	919	ORANG	2014
3	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	LAKI-LAKI	409	ORANG	2014
4	32	JAWA BARAT	3202	KABUPATEN SUKABUMI	PEREMPUAN	295	ORANG	2014
5	32	JAWA BARAT	3203	KABUPATEN CIANJUR	LAKI-LAKI	200	ORANG	2014
...	...	...	...	...	...	...	...	...
432	32	JAWA BARAT	3279	KOTA BANJAR	PEREMPUAN	19	ORANG	2021

Tabel 1. Dataset awal

## 4.3 Data Preparation

Data *preparation* yaitu tahap persiapan data, dilakukan prosedur untuk mempersiapkan data mentah agar dapat digunakan pada tahap pemodelan berikutnya penelitian penyebaran penyakit DBD pada Provinsi Jawa Barat yang akan digunakan saat tahap pemodelan berikutnya.

### 1. Data Selection

data mentah jumlah kasus Demam Berdarah Dengue (DBD) yang didapatkan pada dataset tahun 2014 sampai dengan tahun 2021 dengan banyaknya data 432 data. Kemudian tahap selanjutnya data ini akan diseleksi untuk memudahkan penelitian dan memudahkan analisis, dari seluruhnya 9 atribut saja hanya 3 atribut saja yang dipilih yaitu nama kabupaten kota, jumlah kasus, maupun tahun. Untuk proses pada tahap ini dapat dilihat pada Tabel 2 berikut.

id	nama_kabupaten_kota	jenis_kelamin	jumlah_kasus	tahun
1	KABUPATEN BOGOR	LAKI-LAKI	915	2014
2	KABUPATEN BOGOR	PEREMPUAN	919	2014
3	KABUPATEN SUKABUMI	LAKI-LAKI	409	2014
4	KABUPATEN SUKABUMI	PEREMPUAN	295	2014
5	KABUPATEN CIANJUR	LAKI-LAKI	200	2014
..	..	..	..	..
432	KOTA BANJAR	PEREMPUAN	19	2021

Tabel 2. Data Selection

## 2. Data Preprocessing

Selanjutnya pada tahap kedua ini yaitu data *preprocessing* penelitian ini akan dilakukan persiapan data mentah untuk memudahkan analisis pada penelitian penyebaran penyakit dan virus DBD pada Provinsi Jawa Barat sehingga data siap untuk dijadikan data yang dapat digunakan pada pemodelan selanjutnya. Tahap *preprocessing* pada data adalah tahap di mana data mentah akan diproses melalui proses pembersihan data, salah satunya adalah menangani missing value pada dataset. Namun, pada penelitian ini tidak terdapat missing value, meskipun terdapat data yang berisi 0, namun tidak ada data yang null atau kosong.

Berikut adalah gambar hasil dari data *preprocessing* terdapat, nama\_kabupaten\_kota dengan nilai 0, jumlah\_kasus dengan nilai 0, tahun dengan nilai 0, angka 0 yang berarti pada kolom tersebut masih memiliki nilai namun tidak null atau tidak terisi dengan nilai.

```
[ ] #Data Preprocessing
data_select.isna().sum()

nama_kabupaten_kota    0
jumlah_kasus           0
tahun                  0
dtype: int64
```

Gambar 2. Data Preprocessing

## 3. Data Transformation

Tahap berikutnya yaitu tahap transformasi data atau data *Transformation*, yang dimana data sebelumnya bernilai angka bervariasi maka tahap ini kita ubah menjadi skala 0-1. Transformasi data adalah tahap penting dalam pengembangan pada model data mining dengan melibatkan proses mengubah dataset awal menjadi format yang dapat digunakan untuk analisis lebih lanjut. Salah satu teknik transformasi yang sangat umum digunakan adalah Standard Scaler. Teknik ini untuk membuat rerata data menjadi 0 dan variansi menjadi 1..

```
#Data Transformation atau MinMax Normalization
from sklearn import preprocessing

minmax = preprocessing.MinMaxScaler().fit_transform(X)
df_transform = pd.DataFrame(minmax, index=X.index, columns=X.columns)
df_transform
```

**Gambar 3.** Proses *Data Transformation*

Berdasarkan gambar 3 merupakan beberapa data yang telah di normalisasikan pada skala 0-1 pada hasil data *transformation*

	jumlah_kasus
0	0.394057
1	0.395780
2	0.176141
3	0.127046
4	0.086133
...	...
427	0.151593

**Gambar 4.** Hasil *Data Transformation*

**4.4 Modeling**

Pada tahap pemodelan dalam pengembangan data mining, teknik yang digunakan yaitu teknik clustering yang bertujuan untuk menemukan pola data mining. Algoritma *k-means* ini merupakan algoritma yang umum digunakan, Algoritma ini merupakan teknik clustering berdasarkan pembagian jarak. Sebelum dilakukannya pemodelan, langkah pertama adalah menetapkan cluster. yang bertujuan untuk menetapkan jumlah cluster optimal, pada pemodelan saat ini menggunakan metode *elbow* seperti pada Gambar 5 berikut.

```
from sklearn.cluster import KMeans

clusters = []

for i in range(1,11):
    kmeans = KMeans(n_clusters=i, init='k-means+', random_state=42)
    kmeans.fit(df_transform)
    clusters.append(kmeans.inertia_)
    print('K : '+str(i)+' SSE : '+str(kmeans.inertia_))

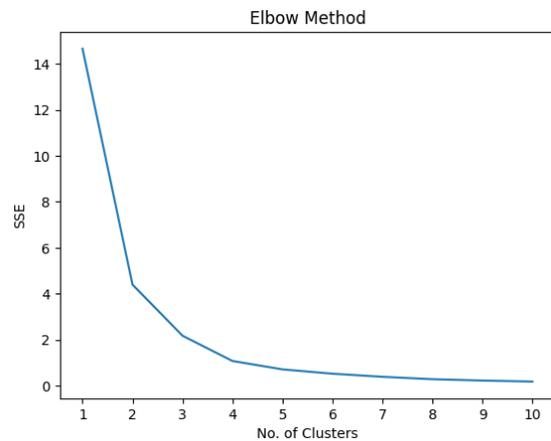
plt.plot(range(1,11),clusters)
plt.xticks(range(1,11))
plt.title('Elbow Method')
plt.xlabel('No. of Clusters')
plt.ylabel('SSE')
plt.show()
```

**Gambar 5.** Source Code Pemodelan *K-Means*

Proses metode *elbow* telah di lakukan maka harus dilakukan visualisasi dalam bentuk grafik untuk mempermudah analisis, grafik metode *elbow* di sajikan pada gambar 6.

Berdasarkan hasil visualisasi metode *elbow* tersebut bahwa cluster optimal yaitu tiga (3) cluster, dengan begitu sudah ditemukan bahwa pada analisis kali ini memiliki jumlah 3 cluster,

Pada gambar 6 merupakan hasil pada metode *elbow*



**Gambar 6.** Grafik Metode *Elbow*

Pada tabel 3 menunjukkan bahwa jumlah anggota *cluster* 0 menunjukkan 139 data dan kriteria rendah 139 data, lalu untuk *cluster* 1 menunjukkan 33 data dengan kriteria tinggi 33 data, dan untuk *cluster* 2 memiliki 260 data dengan kriteria sedang 260 data.

Cluster	Jumlah Anggota Cluster	Jumlah Anggota Kriteria
0	139 Data	Rendah (139 Data)
1	33 Data	Tinggi (33 Data)
2	260 Data	Sedang 260 Data

**Tabel 3.** Jumlah Anggota *Cluster* dan Kriteria

**4.5 Evaluation**

Pada tahap berikut ini yaitu tahap evaluasi. Teknik *silhouette* yang akan digunakan pada tahap ini untuk proses bisa dilihat pada Gambar 7.

```
[ ] from sklearn.metrics import silhouette_score

shil_avg = silhouette_score(df_transform, km.labels_)
print('Nilai Silhouette Coefficient : '+str(shil_avg))
```

**Gambar 7.** Proses *Silhouette Score*

Setelah melakukan proses, maka mendapatkan hasil pada gambar 8. dengan hasil score 0,5931974349414901

Nilai Silhouette Coefficient : 0.5931974349414901

### Gambar 8. Hasil *Silhouette Score*

Pada tahap pengujian dalam pengembangan data mining, dilakukan untuk mengevaluasi kualitas cluster yang dihasilkan oleh algoritma clustering, seperti K-Means Clustering. Evaluasi ini dilakukan untuk mengetahui kedekatan relasi antar objek dan seberapa jauh antar cluster terpisah. Salah satu metode evaluasi yang umum digunakan adalah Silhouette Coefficient. Pada tabel 4 dalam beberapa penelitian, seperti pada analisis cluster dengan menggunakan metode K-Means, ditemukan nilai Silhouette Coefficient yang digunakan untuk mengevaluasi kualitas cluster yang dihasilkan.

Nilai SC	Kualitas	Interpretasi
0,17 – 1,00	Strong	Klaster terbaik sudah ditemukan
0,51 – 0,70	Medium	Penempatan klaster yang wajar
0,26 – 0,50	Weak	Strukturanya lemah, coba metode tambahan
$\leq 0,25$	No Structure	Tidak ada struktur yang ditemukan

Tabel 4. Nilai *Silhouette Score*

### 4.6 Deployment

Pada tahap *Deployment* saat ini akan menerapkan atau dibuat sebuah laporan dari berisi hasil yang sesuai dengan tahapan pada CRISP-DM. Setelah tahap evaluasi ini selesai maka akan implementasikan mode-model yang telah di bangun pada penelitian ini, pada proses deployment ini sudah seluruhnya di implementasikan pada tahap sebelumnya.

### 5. Kesimpulan

Analisis klasifikasi penyebaran pada penyakit DBD di daerah Jawa Barat dengan tingkat penyebaran penyakit menghasilkan 3 cluster uji yaitu cluster 0 dengan 108 data, cluster 1 dengan 33 data, dan cluster 2 dengan 260 data. Pada data yang telah dihasilkan maka Potensi daerah dengan tingkat penyebaran penyakit DBD terbilang tertinggi yang ditujukan kepada

cluster 2 dengan kriteria tinggi maka dari itu indeks terkena penyakit DBD ini pada setiap kota tercatat mencapai 260 data. Berdasarkan hasil penelitian yang dilakukan penulis maka bisa diambil kesimpulan algoritma *K-Means* berhasil dikelompokan yang menghasilkan hasil yang optimal, maka dari itu penelitian ini berhasil untuk memberikan informasi potensi terkena penyakit DBD pada provinsi jawa barat yang terdapat disetiap daerah yang terkena DBD. Maka dari itu kelebihan pada penelitian ini penggunaan nilai rata-rata elbow yang digunakan untuk penentuan cluster optimal dengan 3 anggota cluster optimal dengan begitu 3 clusteri Metode ini dapat memberikan kemudahan dalam proses penelitian, karena dapat menunjukkan jumlah serta nama daerah pada setiap cluster serta hasil. Namun, Terdapat kekurangan pada penelitian ini, yaitu terdapat beberapa data dan anggota cluster yang tidak berada ditempat yang tepat karena pengujian awal yang tidak dapat ditentukan melainkan langsung berdasarkan python sendiri.

### UCAPAN TERIMA KASIH

Terimakasih pada pihak-pihak terkait yang telah memberi dukungan pada penelitian ini. dan tidak lupa terima kasih kepada tuhan yang maha esa

### DAFTAR PUSTAKA

- [1] A. Raksanagara, N. Arisanti, and F. Rinawan, "Dampak Perubahan Iklim Terhadap Kejadian Demam Berdarah Di Jawa-Barat," *J. Sist. Kesehat.*, vol. 1, no. 1, pp. 43–47, 2016, doi: 10.24198/jsk.v1i1.10339.
- [2] Indriany Rahayu, R. Marwati, and D. Rachmatin, "Peramalan Jumlah Penderita DBD di Provinsi Jawa Barat dengan Metode Hybrid Sarimax-Ann," *JMT J. Mat. dan Terap.*, vol. 4, no. 2, pp. 9–19, 2022, doi: 10.21009/jmt.4.2.2.
- [3] H. Prastiwi, J. Pricilia, and E. Raswir, "Implementasi Data Mining Untuk Menentuksn Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering Jurnal Informatika Dan Rekayasa Komputer ( JAKAKOM )," *J. Inform. Dan Rekayasa Komput.*, vol. 1, no. April, pp. 141–148, 2022.
- [4] V. Ramadhan and A. Voutama, "Clustering Menggunakan Algoritma K-Means Pada Penyakit ISPA di Puskesmas Kabupaten Karawang," *J. Pendidik. dan Konseling*, vol. 4, pp. 462–473, 2022.

- [5] M. Syahril, S. Kusnasari, A. Muhazir, and A. Syahputri, "Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Implementasi Data Mining Untuk Rekomendasi Jurusan Menggunakan Algoritma K-Means Clustering Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD," *Tekno. Sist. Inf. dan Sist. Komput. TGD*, vol. 6, pp. 235–245, 2023.
- [6] M. Noperia, Ishak, and V. Winda Sari, "Implementasi Data Mining Pengelompokan Data Nilai Untuk Menentukan Minat Belajar Seni Budaya," *J. Sist. Inf. Tgd*, vol. 2, no. 1, pp. 65–72, 2019, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsi>
- [7] T. P. Yoga, "Optimalisasi K-Means Berbasis Particle Swarm Optimization untuk Hasil Produksi Tanaman Sayuran di Indonesia," vol. 17, 2023.
- [8] A. P. Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *J-SISKO TECH (Jurnal Tekno. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 164, 2023, doi: 10.53513/jsk.v6i1.7351.
- [9] F. Sutomo *et al.*, "Optimization Of The K-Nearest Neighbors Algorithm Using The Optimasi Algoritma K-Nearest Neighbors Menggunakan Metode," vol. 4, no. 1, pp. 125–130, 2023.
- [10] B. Nurseptia, A. Voutama, N. Haeryana, and J. HSRonggo Waluyo, "Penerapan Algoritma K-Means Untuk Mengelompokkan Kabupaten/Kota Dalam Upaya Pemetaan Lapangan Pekerjaan Baru," *J. Tekno. Informasi*, vol. 6, no. 2, pp. 181–186, 2022.
- [11] A. M. M. Fattah, A. Voutama, N. Heryana, and N. Sulistiyowati, "Pengembangan Model Machine Learning Regresi sebagai Web Service untuk Prediksi Harga Pembelian Mobil dengan Metode CRISP-DM," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 5, p. 1669, 2022, doi: 10.30865/jurikom.v9i5.5021.