

**IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK
PENGELOMPOKAN DATA REKAM MEDIS RAWAT JALAN PADA
PKU MUHAMMADIYAH GRABAG**

Amandha Shafa Nabila¹, Maimunah², Endah Ratna Arumi³

Universitas Muhammadiyah Magelang

E-mail: amandanabilla763@gmail.com¹, maimunah@ummgl.ac.id², arumi@ummgl.ac.id³

Abstract

This research aims to manage medical record data at the Muhammadiyah Grabag Community Health Center (PKU) using the K-Means Clustering method. In the digital era, the use of information technology is important for the efficiency of health services, including the management of medical records. Medical record data at PKU Muhammadiyah Grabag continues to increase every day and requires good management to avoid problems such as large storage areas and loss of important data. The K-Means Clustering method is applied to group outpatient data based on the type of treatment received. The dataset used includes data from 2023 to 2024 with a total of 1704 rows. The clustering results show that the data can be grouped into four different clusters, which helps in further analysis and clinical decision making. This research proves that the application of the K-Means Clustering method is effective in managing medical record data, so that it can improve efficiency, quality of service, and support data-based medical research at PKU Muhammadiyah Grabag.

Keywords — Data Mining, Medical Records, K-Means Clustering, PKU Muhammadiyah Grabag, Data Management, Information Technology.

Abstrak

Penelitian ini bertujuan untuk mengelola data rekam medis pada Pusat Kesehatan Umat (PKU) Muhammadiyah Grabag menggunakan metode K-Means Clustering. Dalam era digital, penggunaan teknologi informasi menjadi penting untuk efisiensi pelayanan kesehatan, termasuk dalam pengelolaan rekam medis. Data rekam medis di PKU Muhammadiyah Grabag yang terus bertambah setiap harinya memerlukan manajemen yang baik untuk menghindari masalah seperti tempat penyimpanan besar dan hilangnya data penting. Metode K-Means Clustering diterapkan untuk mengelompokkan data pasien rawat jalan berdasarkan jenis perawatan yang diterima. Dataset yang digunakan mencakup data dari tahun 2023 hingga 2024 dengan total 1704 baris. Hasil clustering menunjukkan bahwa data dapat dikelompokkan menjadi empat kluster yang berbeda, yang membantu dalam analisis lebih lanjut dan pengambilan keputusan klinis. Penelitian ini membuktikan bahwa penerapan metode KMeans Clustering efektif dalam mengelola data rekam medis, sehingga dapat meningkatkan efisiensi, mutu pelayanan, dan mendukung penelitian kedokteran berbasis data di PKU Muhammadiyah Grabag.

Kata Kunci — Data Mining, Rekam Medis, K-Means Clustering, PKU Muhammadiyah Grabag, Pengelolaan Data, Teknologi Informasi.

1. PENDAHULUAN

Di era digital, segalanya menjadi lebih efisien berkat pemanfaatan teknologi informasi. Rekam medis yang dalam bidang kesehatan sering disebut ICD (International Classification of Diseases) adalah rekam medis pasien yang dirawat di suatu rumah sakit atau klinik. Bahasa medis yang digunakan dokter untuk membuat diagnosis dan memberikan tindakan mengenai penyakit ini diterima pasien berupa bahasa medis (rekam medis), yang oleh ahli rekam medis dikodekan menjadi kode ICD. Bahasa kode ini merupakan standar yang harus dibaca sesuai dengan aturan yang berlaku pada ICD, dan dapat digunakan oleh semua dokter, termasuk mereka yang bukan dokter spesialis. Pada akhirnya, rekam medis berbasis teknologi membantu dalam pengembangan pedoman klinis karena dokter dan tenaga kesehatan dapat mengakses data pasien dengan lebih mudah.

Dengan semakin meningkatnya pelayanan kesehatan, pengetahuan kedokteran menjadi semakin penting untuk dipelajari, namun masih banyak masyarakat yang belum memahami sepenuhnya ilmu kedokteran. Rekam medis adalah jenis buku yang memuat data pasien, Data pasien di rumah sakit mengandung informasi yang sangat besar karena setiap pasien yang berkunjung atau memeriksa kesehatannya disana akan dikenakan file medis yang menghasilkan data yang dapat menimbulkan berbagai masalah yang sering terjadi seperti rekam medis yang hilang atau informasi yang perlu diperoleh dan sebagainya. [1]

Salah satu Puskesmas yang ada di Kabupaten Magelang adalah Pusat Kesehatan Umat Muhammadiyah Grabag. Yang memiliki pasien banyak setiap bulannya. Jumlah pasien yang terus berdatangan membuat rekam medis penuh dengan riwayat pasien, dan tidak ada manfaat data ini, meskipun data dapat memberikan informasi berharga kepada peneliti, profesional kesehatan, dan pembuat kebijakan berbasis bukti. Namun, pengelolaan data analisis data yang besar dan kompleks merupakan masalah besar. Penulis mengatasi masalah tersebut dengan data mining menggunakan metode k-means clustering. Selain itu, data mining adalah salah satu teknologi data yang dapat digunakan untuk mengelola data rekam medis, metode clustering k-means juga efektif dalam mengelola data rekam medis. [2][3]

Pusat Kesehatan Masyarakat (PKU) Muhammadiyah Grabag dapat menampung sekitar 1.704 pasien setiap tahunnya, dengan jumlah pasien antara 8 dan 10 per hari dan jumlah pasien yang dapat mencapai 100 hingga 251 per bulan. Data pasien yang dikumpulkan peneliti mencakup sekitar 1.704 pasien pada tahun 2024.

Penerapan metode K-Means Clustering telah banyak dilakukan oleh peneliti sebelumnya, diantaranya tentang penerapan data mining untuk pengelolaan data rekam medis menggunakan metode k-means clustering pada rumah sakit royal prima medan. Kesimpulan yang diperoleh dari analisis data rekam medis di rumah sakit royal prima medan menunjukkan bahwa data dapat dikelompokkan menjadi 4 cluster berdasarkan jenis kelamin dan jenis perawatan, dengan hasil: cluster 1 terdiri dari 18.217 pasien, cluster 2 terdiri dari 1.016 pasien, cluster 3 terdiri dari 396 pasien, dan cluster 4 terdiri dari 307 pasien. [2]

Peneliti lain yaitu tentang analisis data mining untuk clusterisasi data rekam medis menggunakan algoritma k-means pada rumah sakit sylvani binjai. Kesimpulan yang diperoleh yaitu hasil clusterisasi menunjukkan pembagian data menjadi beberapa kelompok berdasarkan jarak terhadap pusat cluster. Pada proses 2 cluster, data terbagi menjadi 346 data dengan diagnosa Myalgia pada laki-laki di Kebun Lada dan 547 data dengan diagnosa Dengue Hemorrhagic Fever pada perempuan di Pahlawan. Proses 3 cluster menghasilkan 277 data dengan diagnosa Hydronephrosis pada laki-laki di Kebun Lada, 319 data dengan diagnosa Asthma pada laki-laki di Pahlawan, dan 297 data dengan diagnosa Congestive Heart Failure pada laki-laki di Kebun Lada. Sedangkan proses 4 cluster membagi data menjadi 268

data dengan diagnosa Hipertension pada perempuan di Kebun Lada, 289 data dengan diagnosa Asthma pada perempuan di Pahlawan, 185 data dengan diagnosa GERD pada laki-laki di Kebun Lada, dan 151 data dengan diagnosa Enlarged Prostat pada laki-laki di Kebun Lada. [3]

Peneliti lain yaitu tentang penerapan data mining untuk pengelompokan data rekam medis pasien berdasarkan jenis penyakit dengan Algoritma K-Means. Kesimpulan yang diperoleh menunjukkan bahwa metode data mining Algoritma K-Means efektif dalam mengelompokkan data rekam medis pasien poli klinik PT. Inecda berdasarkan wilayah, jenis kelamin, dan usia. Pasien terbanyak adalah dewasa (4.912 pasien), diikuti oleh anakanak (1.262 pasien), dan balita (144 pasien). Penyakit terbanyak adalah ISPA dengan 1.985 pasien, disusul oleh penyakit lain-lain seperti jatuh dari motor, kolesterol, dan kontrol kehamilan dengan 2.142 pasien. [4]

Peneliti lain yaitu tentang implementasi algoritma k-means untuk pengelompokan penyakit pasien pada puskesmas warujayeng. Kesimpulannya adalah bahwa pengklasifikasian status kesehatan suatu wilayah masih dilakukan dengan secara manual menggunakan teknik statistik dasar, yang menghasilkan tidak optimal untuk menimbulkan masalah dengan konsistensi data untuk setiap layanan kesehatan. Dari 1000 data, cluster akut terdiri dari 36 item dan non akut terdiri dari 624 item. Ini disebabkan oleh banyaknya data di poliklinik, rumah sakit, puskesmas dan layanan kesehatan. Untuk menemukan informasi tersembunyi dalam data multimediasi dan meningkatkan konsistensi data layanan kesehatan, teknik penambahan data seperti clustering k-means dapat digunakan. [5]

Dengan demikian, untuk meningkatkan efisiensi, kualitas pelayanan, dan penelitian medis di industri kesehatan saat ini, penerapan data mining dengan metode k-means clustering dalam pengelolaan data rekam medis di PKU Muhammadiyah Grabag menjadi sangat penting. Untuk mencapai tujuan ini penulis penerapan data mining dengan metode k-means clustering diterapkan pada rekam medis rawat jalan di PKU Muhammadiyah Grabag. [6]

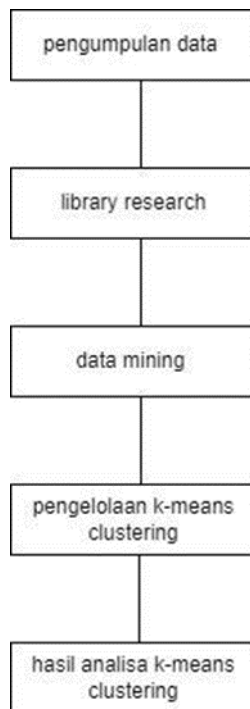
2. METODE

1. Tahapan Penelitian

Dalam menyelesaikan suatu masalah dalam penelitian, peneliti harus memiliki pendekatan atau metode yang tepat untuk memastikan penelitian dapat diselesaikan dengan baik dan menghasilkan hasil yang diharapkan. Metode penelitian digunakan untuk melakukan pencarian secara sistematis dengan memanfaatkan metode ilmiah dan sumber yang relevan. Metode penelitian yang diterapkan dalam studi ini melibatkan pengumpulan informasi melalui riset pustaka (library research). Setelah data terkumpul, langkah berikutnya adalah pengolahan data menggunakan teknik data mining. Setelah data mining, analisis dilanjutkan dengan metode k-means clustering untuk mengidentifikasi jenis perawatan yang sedang berlangsung di PKU Muhammadiyah Grabag. [7]

2. Pengumpulan Data

Untuk mendukung penelitian yang akan dilakukan, penulis mengumpulkan data dari buku dan jurnal penelitian yang relevan. Sumber-sumber ini terkait dengan metode k-means yang digunakan dalam penelitian ini. Dalam penelitian ini, metode library research diterapkan dengan pendekatan kualitatif melalui penelitian lapangan.[8] Dataset yang digunakan berasal dari rekam medis PKU Muhammadiyah Grabag, mencakup data dari tahun 2023 hingga 2024 dengan 8 kolom dan 1704 baris. Dataset ini berisi informasi mengenai pasien rawat jalan. [7][9]



Gambar 1. Tahapan Proses Penelitian

1	No.	No. RM	JK	Umur	Status	Tgl Reg	Cara Bayar	Unit Pelayanan
2	1	00003485	LakiLaki	71 Th	Lama	9/15/23 11:26	BPJS / JKN	PoliSaraf
3	2	00001739	Perempuan	6 Th	Lama	9/20/23 17:35	BPJS / JKN	PoliAnak
4	3	00001896	Perempuan	80 Th	Lama	9/22/23 15:30	BPJS / JKN	PoliSaraf
5	4	00003262	LakiLaki	64 Th	Lama	9/22/23 15:30	BPJS / JKN	PoliSaraf
6	5	00002670	Perempuan	63 Th	Lama	9/22/23 15:38	BPJS / JKN	PoliSaraf
7	6	00003403	Perempuan	15 Th	Lama	9/22/23 15:47	BPJS / JKN	PoliAnak

Gambar 2. Dataset Rawat Jalan 2023-2024

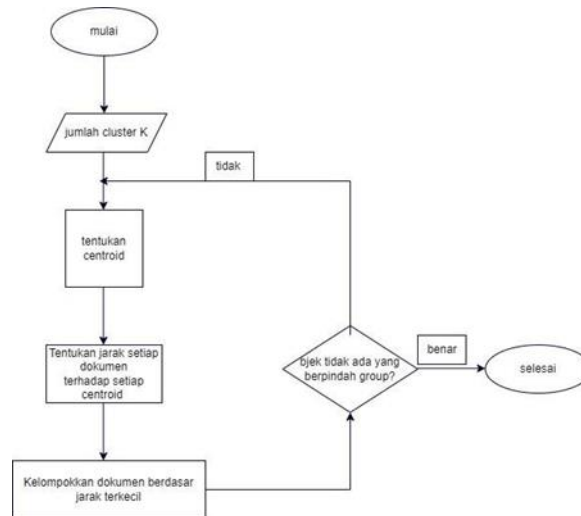
3. Data Mining

Pengumpulan data, ekstraksi, analisis, dan statistik adalah semua bagian dari proses penambangan data. Proses ini juga disebut dengan nama lain seperti penemuan pengetahuan, ekstraksi pengetahuan, analisis data atau pola, dan pencarian informasi.[10]

4. K-Means Clustering

Pada titik ini, ada beberapa langkah yang harus dilakukan dalam proses pengelolaan cluster menggunakan penggabungan K-Means, yaitu:

- Menghitung jumlah kluster K, setelah melakukan preprocecing data dan transformasi data.
- Memilih centroid yang sesuai dengan jumlah K.
- Menghitung jarak objek dan centroid, kemudian mengelompokkan menurut jarak minimumnya.
- Setelah dilakukan pengecekan jika objek bergerak maka proses literasi dilanjutkan.
- Untuk memastikan bahwa objek tidak bergerak, pengecekan dilakukan dan cluster terakhir yang terbentuk dicatat sebagai hasil akhir penggabungan.
- Detail flowchart dengan pengolahan k-means clustering yang menunjukkan kelompokan pada gambar 3.



Gambar 3. Flowchart K-Means Clustering

3. HASIL DAN PEMBAHASAN

1. Identifikasi Data

Data diperoleh dari Pusat Kesehatan Umat (PKU) Muhammadiyah Grabag, mencakup informasi rawat inap pasien dari Februari 2023 hingga Maret 2024. Dataset ini terdiri dari 1704 baris dan 8 kolom.

1	No.	No. RM	JK	Umur	Status	Tgl Reg	Cara Bayar	Unit Pelayanan
2	1	00003485	LakiLaki	71 Th	Lama	9/15/23 11:26	BPJS / JKN	PoliSaraf
3	2	00001739	Perempua n	6 Th	Lama	9/20/23 17:35	BPJS / JKN	PoliAnak
4	3	00001896	Perempua n	80 Th	Lama	9/22/23 15:30	BPJS / JKN	PoliSaraf
5	4	00003262	LakiLaki	64 Th	Lama	9/22/23 15:30	BPJS / JKN	PoliSaraf
6	5	00002670	Perempua n	63 Th	Lama	9/22/23 15:38	BPJS / JKN	PoliSaraf
7	6	00003403	Perempua n	15 Th	Lama	9/22/23 15:47	BPJS / JKN	PoliAnak

Gambar 4. Dataset Pengumpulan Data

1) Library

Beberapa perpustakaan diperlukan untuk pemrosesan data mining dan pemodelan pengelompokan K-means, jadi bab ini membahas tahapan penelitian ini, yaitu import perpustakaan, pengolahan data mining, dan pemodelan pengelompokan K-means.

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder, MinMaxScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import drive
drive.mount('/content/gdrive')
```

Gambar 5. Potongan Kode Library

2. Data Mining

Peneliti akan membahas proses pengelolaan data mining, yaitu:

1) Data Transformation

Transformasi data adalah proses mengubah data dari format atau struktur aslinya ke bentuk yang sesuai untuk analisis lebih lanjut. Dalam konteks clustering menggunakan algoritma k-means di Python, transformasi data melibatkan beberapa langkah penting untuk memastikan data siap digunakan dalam algoritma clustering.[11]

```
# Membuat labelencoder
labelencoder = LabelEncoder()
# Menetapkan nilai numerik dan menyimpannya di kolom lain yang disebut "Jenis Kelamin_Kode"
df["JK"] = labelencoder.fit_transform(df["JK"])
```

Gambar 6. Potongan Kode Data Transformation

```
print(df['JK'])
```

0	0
1	1
2	1
3	0
4	1
...	...
1698	1
1699	1
1700	0
1701	1
1702	1

Name: JK, Length: 1702, dtype: int64

Gambar 7. Gambar Data Setelah Transformation

2) Data Selection

Pemilihan data (data selection) dalam pengolahan k-means clustering dengan Python adalah proses memilih subset data yang paling relevan untuk analisis. Tujuannya adalah meningkatkan efisiensi dan akurasi model clustering dengan fokus pada variabel atau fitur yang paling signifikan dalam membedakan kelompok data. Aspek penting dalam data selection untuk k-means clustering meliputi identifikasi fitur relevan, penghapusan fitur redundan, dan pemilihan sampel yang representatif. Kolom “Unit Pelayanan” berisi data pilihan. Fitur ini memungkinkan peneliti untuk menggunakan datanya sebagai variabel untuk pemodelan dan analisis. [12]

```
features=df["Unit Pelayanan"]
features
```

0	PoliSaraf
1	PoliAnak
2	PoliSaraf
3	PoliSaraf
4	PoliSaraf
...	...
1698	PoliSaraf
1699	PoliSaraf
1700	PoliAnak
1701	PoliSaraf
1702	PoliSaraf

Name: Unit Pelayanan, Length: 1702, dtype: object

Gambar 8. Gambar Potongan Kode Data Selection

3) Membuat Kode Variabel Kategorial

Pengkodean variabel kategorial menggunakan LabelEncoder dalam k-means clustering dengan Python adalah proses mengubah data kategorial menjadi format numerik, di mana setiap kategori unik diberi label angka yang berbeda. LabelEncoder dari library scikit-learn mengonversi setiap nilai kategori menjadi angka unik, yang memungkinkan algoritma k-means menghitung jarak antar data poin dan melakukan clustering. [13]

```
[16] features_encoded = labelencoder.fit_transform(features)
```

Gambar 9. Potongan Kode Pengkodean Variabel Kategorial

4) Melakukan Data Normalization

Data normalization dalam pengolahan k-means clustering dengan Python adalah langkah penting untuk menstandarisasi nilai variabel numerik dalam dataset, sehingga rentang nilai yang berbeda tidak mempengaruhi perhitungan jarak dalam algoritma k-means.

Tujuannya adalah memastikan bahwa hasil clustering yang dihasilkan lebih akurat dan konsisten.

```
[17] scaler = MinMaxScaler()
      features_encoded_resaped = features_encoded.reshape(-1, 1)
      features_scaled = scaler.fit_transform(features_encoded_resaped)

[18] df_scaled = scaler.fit_transform(features_scaled)
```

Gambar 10. Potongan Kode Melakukan Data Normalization

Kode di atas digunakan untuk menskalakan data. Objek `MinMaxScaler()` dibuat dan disimpan dalam variabel pengukur dalam kode ini. Variabel "feature_scaled" adalah variabel di mana skala fitur disimpan. Metode `reshape()` juga mengubah array fitur yang dikodekan dengan "features_encoded" menjadi array dua dimensi dengan ukuran (-1, 1), yang berarti array menjadi satu kolom dan jumlah baris secara otomatis disesuaikan dengan jumlah elemen dalam array. Nilai fitur diubah ke range tertentu, yaitu antara 0 dan 1. Metode penskalaan `MinMaxScaler` mengubah nilai fitur ke range. [14]

3. K-Means Clustering

Penelitian dimulai dengan memodelkan dataset rawat jalan setelah proses data mining selesai.

1) Membuat Kode Penentuan Jumlah Cluster

Pada gambar di bawah ini, menunjukkan kode yang digunakan untuk menentukan nilai objek optimal dalam algoritma k-means clustering, ini juga mencakup perhitungan WSS (Within Cluster Sum of Square) dan skor shilloutte, serta membuat grafik untuk masing masing metrik. Kedua fungsi ini memiliki tujuan untuk membantu dalam menentukan jumlah cluster yang ideal untuk algoritma pembagian k-means. [15]

```
[19] # Menentukan jumlah cluster menggunakan Silhouette Coefficient dan WSS
      silhouette_scores = []
      wss = []

      for n_clusters in range(2, 11):
          kmeans = KMeans(n_clusters=n_clusters, random_state=42, n_init=10) # Set n_init explicitly
          kmeans.fit(df_scaled)

          # Silhouette score
          if len(set(kmeans.labels_)) == 1:
              # Avoid silhouette_score calculation if there's only one cluster
              silhouette_avg = -1
          else:
              silhouette_avg = silhouette_score(df_scaled, kmeans.labels_)
              silhouette_scores.append(silhouette_avg)

          # WSS (Within-Cluster Sum of Square)
          wss.append(kmeans.inertia_)
```

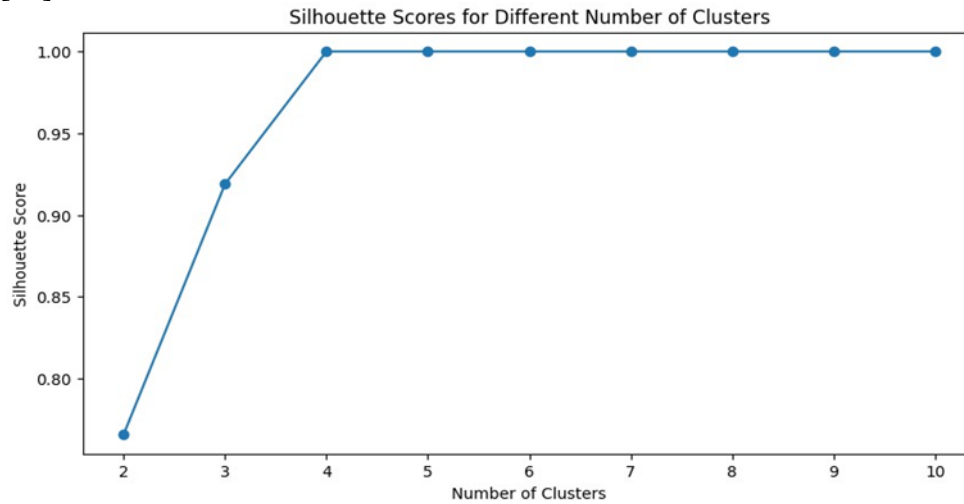
Gambar 11. Potongan Kode Mengidentifikasi Cluster Dengan 2 Metode

```
# Visualisasi Silhouette Scores
plt.figure(figsize=(10, 5))
plt.plot(range(2, 11), silhouette_scores, marker='o')
plt.title('Silhouette Scores for Different Number of Clusters')
plt.xlabel('Number of Clusters')
plt.ylabel('Silhouette Score')
plt.show()

# Visualisasi WSS
plt.figure(figsize=(10, 5))
plt.plot(range(2, 11), wss, marker='o')
plt.title('Within-Cluster Sum of Square (WSS) for Different Number of Clusters')
plt.xlabel('Number of Clusters')
plt.ylabel('WSS')
plt.show()
```

Gambar 12. Potongan Kode untuk Identifikasi Cluster dengan 2 Metode

Setelah clusterisasi K-means selesai, langkah selanjutnya adalah evaluasi, yang biasanya dilakukan dengan metode elbow dan siluet. Kedua metode tersebut menghasilkan grafik yang membantu menentukan jumlah cluster yang ideal untuk suatu kumpulan data. Hasil visualisasi kedua metode setelah menjalankan fungsi evaluasi ditunjukkan pada gambar berikut. [16]



Gambar 13. Visualisasi Hasil Evaluasi Jumlah Cluster Metode Silhouette

Gambar 14. Visualisasi Hasil Evaluasi Jumlah Cluster Metode Elbow

Berdasarkan metode WSS dan skor silhouette, peneliti dapat menyimpulkan bahwa jumlah cluster yang optimal untuk menerapkan k-means clustering adalah 4, yang ditunjukkan oleh lekukan siku pada grafik cluster.

2) Membangun Model K-Means Clustering

Pada langkah ini, peneliti menentukan jumlah cluster yang dimaksud dengan menetapkan nilai $k = 4$ terlebih dahulu. Peneliti juga membuat objek K-Means dengan menggunakan perpustakaan Scikit-Learn. Program diinisialisasi dengan parameter $n_clusters=k$ dan $random_state=42$ untuk mengontrol inisialisasi acak untuk memungkinkan replikasi hasil. Fit dilakukan pada data $k\text{-means.fit}(features_scaled)$ untuk menerapkan algoritma clustering K-means pada data yang telah dinormalisasi. Metode $fit()$ mengidentifikasi pusat cluster terbaik dan mengelompokkan semua data sampel ke dalam cluster yang sesuai berdasarkan metrik jarak geometris atau metrik lainnya.

```
[20] # Membuat Model K-Means Clustering dengan jumlah kluster yang optimal (misalkan 3)
      optimal_clusters = 4
      kmeans = KMeans(n_clusters=optimal_clusters, random_state=42, n_init=10) # Set n_init explicitly
      df['Cluster'] = kmeans.fit_predict(df_scaled)
```

Gambar 15. Potongan Kode Membuat Model Sesuai Jumlah Cluster

Model K-Means akan mengandung pusat cluster yang ditemukan dan inersia setelah proses fitting selesai (SSE, Sum of Square Errors). Untuk menganalisa hasil clustering atau memprediksi cluster data baru, model tersebut dapat digunakan. [17]

3) Membuat Centroid

Peneliti kemudian akan menemukan centroid, yang akan digunakan untuk menemukan koordinat pusat kelompok saat membuat model K-means. Untuk mendapatkan indeks dan koordinat pusat cluster_centers, gunakan fungsi $enumerate()$ untuk menampilkan hasil clustering.

```
[21] # Membuat Centroid
centroids = kmeans.cluster_centers_

[22] cluster_labels = kmeans.labels_

#Cetak label klaster untuk setiap sampel
#for i, label in enumerate (cluster_labels):
# print("Sampel", i, "Ditugaskan Ke Klaster", label)

#Cetak koordinat pusat klaster
for i, center in enumerate(centroids):
    print("Pusat Cluster", i, ":", center)
```

Gambar 16. Potongan Kode Membuat Centroid

Setelah program selesai, peneliti dapat melihat nilai centroid, yang ditunjukkan pada gambar berikut:

```
Pusat Cluster 0 : [0.66666667]
Pusat Cluster 1 : [0.33333333]
Pusat Cluster 2 : [1.]
Pusat Cluster 3 : [-3.33066907e-16]
```

Gambar 17. Hasil Nilai Centroid

Hasil clustering menunjukkan bahwa setiap centroid merupakan representasi dari satu cluster, dan koordinat centroid menunjukkan pusat atau representasi rata-rata dari sampel yang termasuk dalam cluster tersebut.

4) Membuat Kode Penentuan Jarak setiap Data Centroid

Setelah menentukan centroid, peneliti kemudian menghitung jarak data ke centroid, yang ditunjukkan dalam cuplikan kode berikut:

```
[53] # Menghitung Jarak Setiap Titik Data
# Data yang akan digunakan
X = np.array([[1, 2], [1, 4], [1, 0], [4, 2], [4, 4], [4, 0]])
# Menghitung jarak antara setiap titik data dengan masing-masing pusat klaster
distances = []
for i, data_point in enumerate(X):
    cluster_distances = []
    for centroid in centroids:
        distance = np.linalg.norm(data_point - centroid)
        cluster_distances.append(distance)
    distances.append(cluster_distances)

# Menampilkan hasil jarak
for i, data_point in enumerate(X):
    print("Jarak data", i+1, "ke masing-masing pusat klaster:")
    for j, distance in enumerate(distances[i]):
        print("Klaster", j+1, ":", distance)
```

Gambar 18. Menentukan Jarak Setiap Data Centroid

Gambar dibawah ini akan menunjukkan hasil perhitungan jarak setiap data cluster terhadap setiap centroid:

```

Jarak data 1 ke masing-masing pusat klaster:
Klaster 1 : 1.3743685418725538
Klaster 2 : 1.795054935711501
Klaster 3 : 1.0000000000000004
Klaster 4 : 2.2360679774997902
Jarak data 2 ke masing-masing pusat klaster:
Klaster 1 : 3.3499585403736303
Klaster 2 : 3.7267799624996494
Klaster 3 : 3.0000000000000004
Klaster 4 : 4.123105625617661
Jarak data 3 ke masing-masing pusat klaster:
Klaster 1 : 0.7453559924999298
Klaster 2 : 0.7453559924999299
Klaster 3 : 0.9999999999999996
Klaster 4 : 1.0000000000000004
Jarak data 4 ke masing-masing pusat klaster:
Klaster 1 : 3.590109871423003
Klaster 2 : 4.0276819911981905
Klaster 3 : 3.16227766016838
Klaster 4 : 4.47213595499958
Jarak data 5 ke masing-masing pusat klaster:
Klaster 1 : 4.714045207910317
Klaster 2 : 5.185449728701348
Klaster 3 : 4.242640687119286
Klaster 4 : 5.656854249492381
Jarak data 6 ke masing-masing pusat klaster:
Klaster 1 : 3.39934634239519
Klaster 2 : 3.6817870057290865
Klaster 3 : 3.16227766016838
Klaster 4 : 4.0

```

Gambar 19. Hasil Dari Jarak Setiap Data Centroid

Sebagai hasil dari jarak data ke masing-masing pusat klaster yang diberikan, dapat disimpulkan bahwa:

Jarak Data 1:

Cluster 1 memiliki jarak terdekat dengan cluster 3 dengan 1.0000000000000004, cluster 4 dengan 2.2360679774997902, dan cluster 1 memiliki jarak rata-rata dari cluster 1 dengan 1.3743685418725538. Cluster 2 memiliki jarak rata-rata dari cluster 1 dengan 1.795054935711501.

Jarak Data 2:

Cluster 2 memiliki jarak terdekat dengan Cluster 3 dengan 3.0000000000000004. Cluster 4 memiliki jarak terbesar dengan 4.123105625617661. Cluster 1 memiliki jarak rata-rata sebesar 3.3499585403736303. Cluster 2 memiliki jarak rata-rata sebesar 3.7267799624996494.

Jarak Data 3:

Cluster 3 berada di jarak terdekat dengan Cluster 1 dengan jarak 0.7453559924999298 dan 0.7453559924999298, masing-masing. Cluster 4 berada di jarak terjauh dari Data 3, dengan jarak 1.0000000000000004. Cluster 3 berada di jarak rata-rata dari Cluster 3, dengan jarak 0,9999999999999996.

Jarak Data 4:

Kelompok 4 terdekat dengan cluster 3 dengan 3.16227766016838. Kelompok 4 terjauh dari cluster 4 dengan 4.47213595499958. Kelompok 4 memiliki jarak rata-rata dari cluster 1 dengan 3.590109871423003. Kelompok 2 memiliki jarak rata-rata dari cluster 4 dengan 4.0276819911981905.

Jarak Data 5:

Cluster 5 terdekat dengan cluster 3 dengan 4.242640687119286; cluster 4 terjauh dengan 5.656854249492381; dan cluster 1 terjauh dengan 4.714045207910317. Cluster 2 terjauh dengan 5.185449728701348.

Jarak Data 6:

Data 6 berada di jarak terdekat dengan kelompok 3, yaitu 3.16227766016838. Data 6 berada di jarak terjauh dari kelompok 4, yaitu 4,0. Data 6 memiliki jarak rata-rata ke kelompok 1, yaitu 3.39934634239519, dan ke kelompok 2 yaitu 3.6817870057290865.

Dengan demikian, berdasarkan jarak terdekat, Data 1, Data 2, Data 4, dan Data 6 dapat diklasifikasikan sebagai bagian dari klaster 3. Data 3 dapat diklasifikasikan sebagai bagian dari klaster 1 atau 2, dan Data 5 juga sebagai bagian dari klaster 3. Hal ini menunjukkan

bahwa beberapa data memiliki kedekatan yang signifikan dengan pusat kluster tertentu, namun tidak ada yang berpindah jarak group, sehingga penelitian dengan menggunakan algoritma metode k-means clustering sudah selesai. [18]

5) Hasil Analisis Visualisasi dan Pembagian Data Cluster

Setelah melakukan clustering menggunakan metode K-means, peneliti menganalisis, memahami, dan memberikan penjelasan tentang pemetaan sebaran data pada setiap cluster.

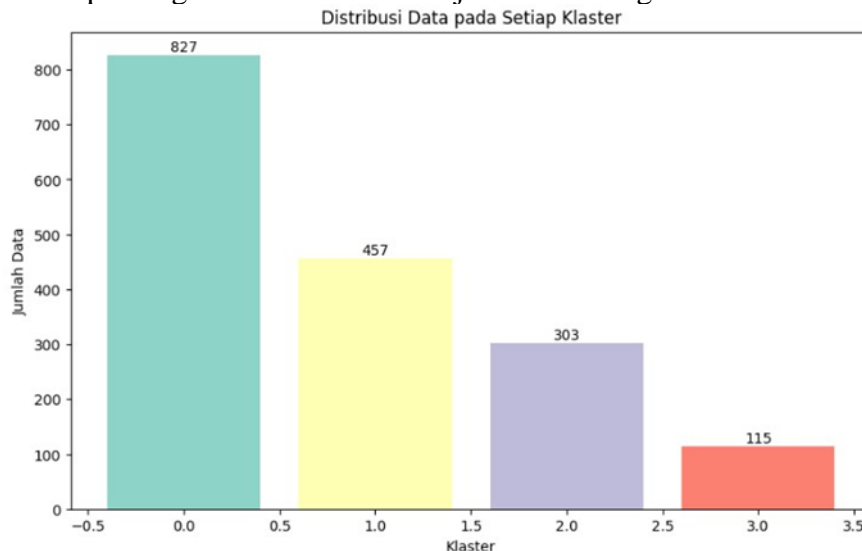
```
[56] # Membuat Visualisasi Hasil Kluster Yang Telah Terbentuk
colors = sns.color_palette("Set3", n_colors=optimal_clusters)
unique_clusters, cluster_counts = np.unique(cluster_labels, return_counts=True)

plt.figure(figsize=(10, 6))
plt.bar(unique_clusters, cluster_counts, color=colors)
plt.xlabel('Kluster')
plt.ylabel('Jumlah Data')
plt.title('Distribusi Data pada Setiap Kluster')

# Menambahkan label jumlah data pada setiap bar
for i in range(len(unique_clusters)):
    plt.text(unique_clusters[i], cluster_counts[i], str(cluster_counts[i]), ha='center', va='bottom')

plt.show()
```

Gambar 20. Potongan Kode untuk Membuat Visualisasi Data Cluster Hasil analisis pembagian data cluster ditunjukkan dalam gambar berikut:



Gambar 21. Hasil Visualisasi Pembagian Data Setiap Cluster

Gambar di atas menunjukkan hasil pemisahan data dari setiap cluster. 827 (48,5%) dari 1704 data pasien di Cluster 1, 457 (26,8%) di Cluster 2, 303 (17,8%) di Cluster 3 dan 115 (6,75%) di Cluster 4. [2][19]

6) Tabel Hasil Analisis Data setiap Cluster

Tabel berikut menunjukkan hasil akhir berdasarkan data:

Tabel 1. Hasil Analisa Data Cluster 1 dan 2

Hasil dari Analisa Cluster 1	Hasil dari Analisa Cluster 2
Dari 1704 data pasien rawat jalan, terdapat sebanyak 827 orang yang berasal dari unit pelayanan poli saraf.	Dari 1704 data pasien rawat jalan, terdapat sebanyak 457 orang yang berasal dari unit pelayanan poli dalam/interma.

Jenis Kelamin :	Jenis Kelamin :
1. Perempuan : 483	1. Perempuan : 265
2. Laki Laki : 344	2. Laki Laki : 192

Tabel 2. Menunjukkan Hasil Analisa Data Cluster 3 dan 4

Hasil dari Analisa Cluster 3	Hasil dari Analisa Cluster 4
Dari 1704 data pasien rawat jalan, terdapat sebanyak 303 orang yang berasal dari unit pelayanan poli umum.	Dari 1704 data pasien rawat jalan, terdapat sebanyak 115 orang yang berasal dari unit pelayanan poli anak.
Jenis Kelamin : 1. Perempuan : 161 2. Laki Laki : 142	Jenis Kelamin : 1. Perempuan : 56 2. Laki Laki : 59

Dari tabel di atas, kita dapat menemukan empat kolom cluster. Berdasarkan analisis Cluster, data pasien rawat jalan di Pusat Kesehatan Umat (PKU) Muhammadiyah Grabag dapat disimpulkan sebagai berikut: Cluster 1: Dari 1704 data pasien, 827 orang berasal dari unit pelayanan poli saraf dengan distribusi jenis kelamin 483 perempuan dan 344 laki-laki. Cluster 2: Dari 1704 data pasien, 457 orang berasal dari unit pelayanan poli dalam/interma dengan distribusi jenis kelamin 265 perempuan dan 192 laki-laki. Cluster 3 : Dari 1704 data pasien, 303 orang berasal dari unit pelayanan poli umum dengan distribusi jenis kelamin 161 perempuan dan 142 laki-laki. Cluster 4: Dari 1704 data pasien, 115 orang berasal dari unit pelayanan poli anak dengan distribusi jenis kelamin 56 perempuan dan 59 laki-laki. [2][20]

3. KESIMPULAN

Peneliti dapat membuat kesimpulan berikut dari penelitian yang dilakukan: metode clustering K-means digunakan untuk mengelola data rekam medis di Puskesmas Muhammadiyah Grabag. Hasil analisis menunjukkan bahwa data rekam medis di Pusat Kesehatan Masyarakat (PKU) Muhammadiyah Grabag dapat dibagi menjadi empat kelompok dengan dua variabel, yaitu jenis kelamin dan unit pelayanan. Kelompok 1 terdiri dari 827 pasien, menjadi 457 pasien, kelompok 2 terdiri dari 303 pasien, dan kelompok 4 terdiri dari 115 pasien rawat jalan.

DAFTAR PUSTAKA

- [1] S. M. Putri¹, U. Hayati², and R. Dzulkarnaen³, "Perancangan Arsitektur Electronic Medical Record (EMR) menggunakan Metode Enterprise Architecture Planning (EAP) Arsitektur Enterprise Jurnal JOINT STMIK 'AMIKBANDUNG'".
- [2] W. Purba et al., "PENERAPAN DATA MINING UNTUK PENGELOLAAN DATA REKAM MEDIS MENGGUNAKAN METODE K-MEANS CLUSTERING PADA RUMAH SAKIT ROYAL PRIMA MEDAN," J. TEKINKOM, vol. 6, no. 1, 2023, doi: 10.37600/tekinkom.v6i1.857.
- [3] W. Surya Nanda, A. M. H. Pardede, M. Simanjuntak, F. T. Informatika, and S. Kaputama, "ANALISIS DATA MINING UNTUK KLASTERISASI DATA REKAM MEDIS MENGGUNAKAN ALGORITMA K-MEANS PADA RUMAH SAKIT SYLVANI BINJAI," Indones. J. Educ. Comput. Sci., vol. 1, no. 3, p. 2023.
- [4] R. Ordila, R. Wahyuni, Y. Irawan, and M. Yulia Sari, "PENERAPAN DATA MINING

UNTUK PENGELOMPOKAN DATA REKAM MEDIS PASIEN BERDASARKAN JENIS PENYAKIT

DENGAN ALGORITMA CLUSTERING (Studi Kasus : Poli Klinik PT.Inecda),” *J. Ilmu Komput.*, vol. 9, no. 2, pp. 148–153, 2020, doi: 10.33060/jik/2020/vol9.iss2.181.

- [5] U. Ma'rifatin, “Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien Pada Puskesmas Warujayeng,” *Semin. Nas. Inov. Teknol.*, no. 2549–7952, pp. 285–291, 2020.
- [6] N. H. Harani, C. Prianto, and F. A. Nugraha, “Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python,” *J. Manaj. Inform.*, vol. 10, no. 2, pp. 133–146, 2020, doi: 10.34010/jamika.v10i2.2683.
- [7] U. Hasanah, A. Hendra Azhar, and I. Artikel, “Data Mining Menentukan Balita Yang Mendapatkan Makanan Tambahan Dari Kemenkes Dengan Metode Naive Bayes (Studi Kasus: Puskesmas Pembantu Bantan),” *J. Rekayasa Sist.*, vol. 1, no. Mei, p. 546, 2023, [Online]. Available: <https://www.doi.org/10.22303/upu.1.1.2021.01-10>
- [8] D. O. Dacwanda and Y. Nataliani, “Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan,” *Aiti*, vol. 18, no. 2, pp. 125–138, 2021, doi: 10.24246/aiti.v18i2.125-138.
- [9] J. Khomsiyah, A. Ramdhani, A. F. Damayanti, and D. Rohman, “Penerapan Algoritma K-means Clustering untuk Pengelompokan Wilayah Rawan Banjir,” *J. Ilm. Betrik*, vol. 12, no. 3, pp. 249–253, 2021, [Online]. Available: <https://www.neliti.com/publications/460035/penerapan-algoritma-k-meansclustering-untuk-pengelompokan-wilayah-rawan-banjir#cite>
- [10] D. Andreswari, R. Efendi, and K. Prastio, “Clustering Data Rekam Medis untuk Penentuan Penyakit Endemi di Daerah Kabupaten Bengkulu Selatan dengan Mengimplementasikan Metode FUZZY C-MEANS,” *J. Rekursif*, vol. 11, no. 1, pp. 42–52, 2023, [Online]. Available: <http://ejournal.unib.ac.id/index.php/rekursif/42>
- [11] A. A. I. I. P. Nengah Widya Utami, “Penerapan Data Mining Untuk Mengetahui Pola Pemilihan Program Studi Di Stmik Primakara Menggunakan Algoritma K-Means ...,” *J. Teknol. Inf. dan ...*, vol. 3, pp. 456–463, 2021, [Online]. Available: <http://jurnal.undhirabali.ac.id/index.php/jutik/article/view/1540>
- [12] F. Handayani, “Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar,” *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.6733.
- [13] A. Badruttamam, S. Sudarno, and D. A. I. Maruddani, “PENERAPAN ANALISIS KLAS TER K-MODES DENGAN VALIDASI DAVIES BOULDIN INDEX DALAM MENENTUKAN KARAKTERISTIK KANAL YOUTUBE DI INDONESIA (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade),” *J. Gaussian*, vol. 9, no. 3, pp. 263–272, 2020, doi: 10.14710/j.gauss.v9i3.28907.
- [14] Y. P. Sari, A. Primajaya, and A. S. Y. Irawan, “Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang,” *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, p. 229, 2020, doi: 10.35314/isi.v5i2.1457.
- [15] E. A. Saputra and Y. Nataliani, “Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means,” *J. Inf. Syst. Informatics*, vol. 3, no. 3, pp. 424–439, 2021, doi: 10.51519/journalisi.v3i3.164.
- [16] W. Hardi, E. Wisnu Ananta Kusuma, and S. Basuki, “Pengelompokan Topik Dokumen Berbasis Text Mining Dengan Algoritme K-Means: Studi Kasus Pada Dokumen Kedutaan Besar Australia Jakarta,” *VISI PUSTAKA Bul. Jar. Inf. Antar Perpust.*, vol. 21, no. 1, pp. 67–76, 2019, [Online]. Available: <https://journals.web.id/vp/article/view/77>

- [17] Remonaldi Purba, "Grouping Medical Record Data By Type Diseases With K-Means Algorithm," *J. Artif. Intell. Eng. Appl.*, vol. 1, no. 2, pp. 128–134, 2022, doi: 10.59934/jaiea.v1i2.78.
- [18] B. Hartono, S. Eniyati, and K. Hadiono, "Perbandingan Metode Perhitungan Jarak pada Nilai Centroid dan Pengelompokan Data Menggunakan K-Means Clustering," *J. Sist. Komput. dan Inform.*, vol. 4, no. 3, p. 503, 2023, doi: 10.30865/json.v4i3.6021.
- [19] T. Hardiani, "Analisis Clustering Kasus Covid 19 di Indonesia Menggunakan Algoritma K-Means," *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 2, pp. 156–165, 2022, doi: 10.23887/janapati.v11i2.45376.
- [20] Ainur Rahman and H. Suroyo, "Analisis Data Produk Elektronik Di E-Commerce Dengan Metode Algoritma K-Means Menggunakan Python," *J. Adv. Inf. Ind. Technol.*, vol. 3, no. 2, pp. 11–18, 2021, doi: 10.52435/jaiit.v3i2.158.