

PENERAPAN DATA MINING UNTUK PENGELOLAAN DATA REKAM MEDIS MENGUNAKAN METODE K-MEANS CLUSTERING PADA RUMAH SAKIT ROYAL PRIMA MEDAN

Windania Purba¹⁾, Gamaliel Armando Sembiring²⁾, Andreas Saputra³⁾,
Mawar Theresia Turnip⁴⁾, Ben Jua Ivand Manihuruk⁵⁾

^{1,2,3,4,5} Fakultas Sains Dan Teknologi, Universitas Prima Indonesia

windania@unprimdn.ac.id, gamaliel.sembiring11@gmail.com, tniasinaga@gmail.com,
mawarturnip11@gmail.com, benjuaivand05@gmail.com

Abstract



In this digital era, medical record data in hospitals has grown to be very large and complex. This medical record data includes information about the patient, diagnosis, treatment, and other medical history. Efficient and effective management of medical record data is essential to improve the quality of health services, appropriate decision-making, and medical research. This study uses data mining techniques with the K-Means Clustering method to cluster patient medical record data. Cluster 1 consists of 1827 people suffering from Emergency, Orthopedics, Obgyn, Internal Medicine, Pulmonary, NICU/PISU, Heart Disease, Perinatology, Neonatal and Growth and Development, Obstetrics Oncology, as well as male and female 9227 and 8990 respectively. Cluster 4 consists of 417 people who suffer from Urology, ENT, General, Neurology, Rheumatology diseases, and the male gender is 195 people and the female gender is 112 people. by using data mining, researchers can find new information about how royal prima medan hospital manages various types of care. Researchers hope to be a reference for hospitals, to be able to socialize and prevent sources of disease based on gender and treatment.

Keywords: Cluster, Data Mining, K-Means Clustering, Medical Record

1. PENDAHULUAN

Di masa digital disaat ini, data rekam medis rumah sakit sangat besar dan kompleks dengan mencakup data tentang penderita diagnosis, pengobatan, serta riwayat medis yang ada. Pengelolaan data rekam medis yang baik sangat berguna buat pengambilan keputusan yang tepat dan penelitian medis akan berguna untuk meningkatkan layanan kesehatan.

Akan tetapi pengelolaan data rekam medis yang baik masih menjadi tantangan untuk rumah sakit. Pengelolaan data rekam medis bagian penting dari pengelolaan rumah sakit karena data ini merupakan informasi penting dalam memberikan layanan kesehatan kepada pasien [1].

Oleh sebab itu, pengelolaan data rekam medis wajib dilakukan dengan baik serta efektif dan salah satu teknologi data yang bisa digunakan

buat pengelolaan data rekam medis merupakan dengan memanfaatkan *data mining*.

Data mining menggambarkan suatu proses yang memakai cara statistik, matematika, kecerdasan buatan, serta mesin pembelajaran yang mengekstraksi serta mengenali data yang berguna dan pengetahuan yang terkait dari berbagai data yang sangat besar [2]–[13]. Salah satu guna *data mining* merupakan *k-means clustering*. Dalam konteks ini, penerapan *data mining* dengan menggunakan metode *k-means clustering* dapat menjadi solusi yang efektif untuk mengelola data rekam medis di rumah sakit. Metode *k-means clustering* menggambarkan proses pengelompokan titik-titik informasi ke dalam 2 kelompok ataupun lebih sehingga titik-titik informasi yang tercantum di dalam kelompok yang sama lebih mirip satu sama lain [14].

Penelitian terdahulu terkait dengan topik penelitian ini yaitu sebagai berikut: penggunaan *data mining* dengan metode *k-means clustering*, yang dapat mengidentifikasi informasi rekam medis bersumber pada jenis kelamin, serta usia [16]. Dengan adanya *data mining* prosedur *k-means clustering*, dapat mengenali data rekam medis dari rumah sakit anwar medika medis berdasarkan kecamatan, diagnosa penyakit, umur serta jenis kelamin [2].

Dengan demikian, penerapan *data mining* dengan memanfaatkan metode *k-means clustering* dalam mengelola data rekam medis pada Rumah Sakit Royal Prima Medan mempunyai relevansi yang besar dalam menaikkan efisiensi, mutu pelayanan, serta penelitian kedokteran di industri kesehatan dikala ini.

Berlandaskan latar belakang diatas yang maka diterapkan penggunaan *data mining* dengan menggunakan metode *k-means clustering* dengan memanfaatkan data rekam medis pada Rumah Sakit Royal Prima Medan, yang bertujuan untuk mendapatkan informasi guna mengelola bermacam jenis perawatan bagi pasien yang sedang dirawat pada Rumah Sakit Royal Prima Medan sebagai pedoman bagi rumah sakit guna meningkatkan mutu pelayanan dalam penangkalan sumber penyakit.

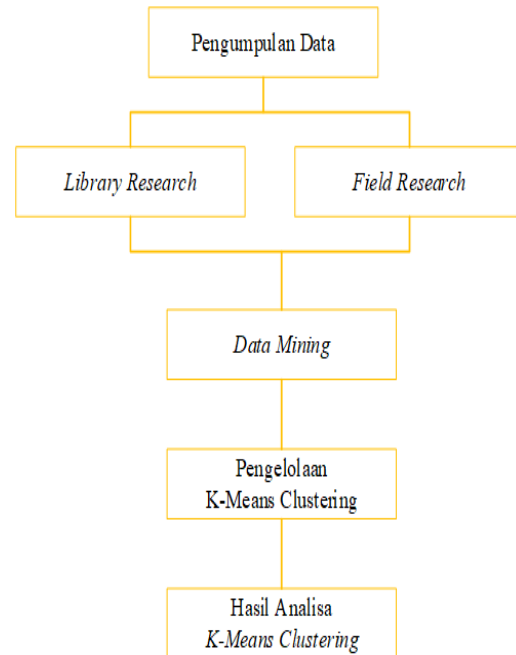
2. METODE PENELITIAN

Metode penelitian yang digunakan dalam menyelesaikan penelitian adalah dengan mengumpulkan informasi dari *library research* serta *field research*, setelah dikumpulkan dilanjutkan dengan Pengelolaan *data mining*, sesudah melakukan proses pengelolaan *data mining* berikutnya dilanjutkan dengan menggunakan *k-means clustering* untuk mencari informasi jenis perawatan yang sedang dirawat pada Rumah Sakit Royal Prima Medan. Proses penelitian terdapat pada gambar 1.

2.1. Pengumpulan Data

Peneliti mengumpulkan informasi serta data lewat *library research* serta *field research*. *Library Research* merupakan aktivitas penelitian

yang dilakukan dengan mengumpulkan data serta informasi yang terdapat di perpustakaan ataupun sumber dari internet yang berkaitan dengan permasalahan yang akan dipecahkan [15].



Gambar 1. Proses Penelitian

Metode pengamatan (*field research*) pada penelitian ini menggunakan penelitian lapangan dengan pendekatan kualitatif [17]. Pada penelitian ini *dataset* yang digunakan berasal dari data rekam medis Rumah Sakit Royal Prima Medan. *Dataset* ini diambil dari tahun 2021 hingga 2022, data ini mempunyai 5 kolom serta 19936 baris. Yang dimana data ini berisikan informasi data pasien yang tengah dirawat. Detail *dataset* tersebut bisa dilihat pada gambar 2.

	A	B	C	D	E
1	no_pendaftaran	jeniskelamin	jeniskasuspeny	tgl_pendaftaran	ruangan_nama
2	RJ2105030052	PEREMPUAN	Bedah	03/05/2021 9:27	Lantai 15
3	RJ2102170209	PEREMPUAN	Bedah	17/02/2021 15:55	Lantai 9
4	RJ2101150060	LAKI-LAKI	Bedah	15/01/2021 10:36	Lantai 9
5	RJ2205300032	LAKI-LAKI	Bedah Digestif	30/05/2022 9:40	B-Lantai 12
6	RJ2205260004	LAKI-LAKI	Bedah Digestif	26/05/2022 13:26	Lantai 12
7	RJ2205230020	LAKI-LAKI	Bedah Digestif	23/05/2022 9:15	B-Lantai 11
8	RJ2205190007	PEREMPUAN	Bedah Digestif	19/05/2022 8:33	B-Lantai 5
9	RJ2205160009	PEREMPUAN	Bedah Digestif	16/05/2022 10:16	B-Lantai 5
10	RJ2204260034	LAKI-LAKI	Bedah Digestif	26/04/2022 9:48	B-Lantai 5

Gambar 2. *Dataset* Rawat Inap 2021 – 2022

2.2. Data Mining

Proses *data mining* merupakan proses terdiri pengumpulan data, ekstraksi data, analisa data, dan statistik data [18]–[22]. Proses ini juga umum dikenal sebagai *knowledge discovery*, *knowledge extraction*, *data/pattern analysis*, *information harvesting*, dan lainnya.

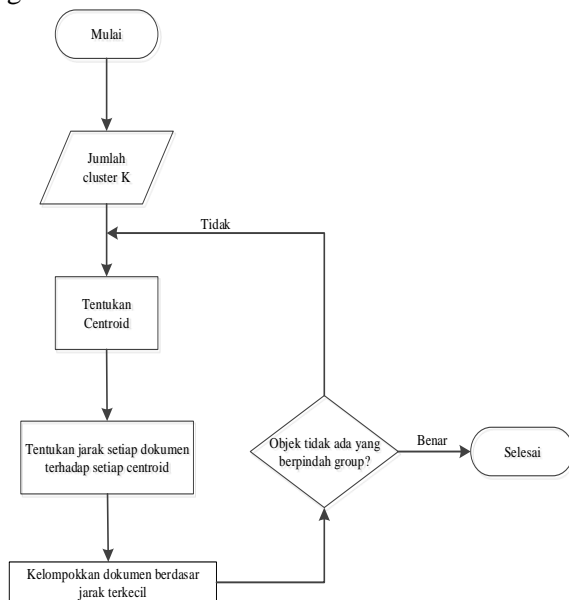
2.3. K-Means Clustering

Pada tahap ini, terdapat beberapa tahap yang perlu dilakukan untuk proses pengelolaan *cluster k-means clustering*, yaitu [17] :

- Melakukan *preprocessing* data dilanjutkan transformasi data, kemudian menentukan jumlah *cluster* (*Number of K*)
- Memilih *centroid* sebanyak random sebanyak jumlah *K* yang ditentukan,
- Menghitung jarak *centroid* ke objek dan mengelompokkan berdasarkan jarak minimum,
- Melakukan pengecekan, jika objek bergerak maka dilanjutkan proses iterasi,
- Jika objek tidak bergerak maka *cluster* terakhir dicatat sebagai hasil *cluster* yang terbentuk.

f.

Detail *flowchart* dari *k-means clustering* pada gambar 3.



Gambar 3. *Flowchart K-Means Clustering*

3. HASIL DAN PEMBAHASAN

Permasalahan yang dialami dalam penelitian ini merupakan mengumpulkan informasi dari data rekam medis yang sudah terkumpul dan buat mengenali kelompok pengidap dengan karakteristik yang sama. Oleh sebab itu, rumusan permasalahan dalam penelitian ini merupakan bagaimana menerapkan *data mining* dengan memanfaatkan metode *k-means clustering* dengan menggunakan data rekam medis pada Rumah Sakit Royal Prima Medan, sehingga dapat digunakan sebagai pedoman buat rumah sakit guna meningkatkan kualitas pelayanan dalam penangkalan sumber penyakit.

3.1. Identifikasi Data

Data diambil melalui Rumah Sakit Royal Prima Medan, informasi ini merupakan data rawat inap pasien dari tahun Januari - 2021 hingga Juni - 2022. Dimana tiap informasi tersebut mempunyai jumlah baris sebanyak 19936 dengan 5 kolom.

no_pendaftaran	jeniskelamin	jeniskasuspenyakit_nama	tgl_pendaftaran	ruangan_nama
RJ2105030052	PEREMPUAN	Bedah	03/05/2021 09:27	Lantai 15
RJ2102170209	PEREMPUAN	Bedah	17/02/2021 15:52	Lantai 9
RJ2101150060	LAKI-LAKI	Bedah	15/01/2021 10:39	Lantai 9
RJ2205300032	LAKI-LAKI	Bedah Digestif	30/05/2022 09:40	B-Lantai 12
RJ2205260004	LAKI-LAKI	Bedah Digestif	26/05/2022 13:28	Lantai 12
RJ2205230020	LAKI-LAKI	Bedah Digestif	23/05/2022 09:15	B-Lantai 11
RJ2205190007	PEREMPUAN	Bedah Digestif	19/05/2022 08:33	B-Lantai 5
RJ2205160009	PEREMPUAN	Bedah Digestif	16/05/2022 10:18	B-Lantai 5
RJ2204260034	LAKI-LAKI	Bedah Digestif	26/04/2022 09:48	B-Lantai 5

Gambar 4. Dataset Pengumpulan Data

3.2. Pengolahan Data

Pada sub bab ini mangulas tentang tahapan penelitian adalah dari import Libraries, proses pengolahan *data mining* serta *k-means clustering*.

3.3. Library

Beberapa libraries ini diperlukan untuk melakukan pengolahan *data mining*, dan melakukan pemodelan *k-means clustering*.

```
import pandas as pd
import re
import string
import math
import numpy as np
import seaborn as sns
sns.set_style('whitegrid')
import matplotlib.pyplot as plt
import math
import warnings
warnings.filterwarnings("ignore")

from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import silhouette_score
from sklearn.metrics import silhouette_samples, silhouette_score
from sklearn.cluster import KMeans
```

Gambar 5. Potongan Kode Libraries Yang Dipakai Untuk Penelitian

3.4. Data Mining

Peneliti akan membahas proses pengolahan data mining yang sudah dijelaskan sebelumnya, yaitu :

3.4.1. Hasil Data Cleaning

Data cleaning berguna buat menghapus *noise* data yang bisa mengusik proses penelitian [23].

```
# MENGAKSES DATA DARI GOOGLE DRIVE KE GOOGLE COLAB
df = pd.read_csv("content/drive/MyDrive/AA/Data Pasien Rawat Inap 2021 - 2022.csv", header=0, delimiter=";")
# Mengecek jumlah nilai yang hilang (missing values) pada setiap kolom
df[["Tanggal Pendaftaran", "Jenis Kelamin", "Jenis Perawatan"]].isnull().sum()
# MENGHAPUS KOLON DATASET YANG BERMASALAH
df.dropna(inplace=True)
```

Gambar 6. Potongan Kode *Data Cleaning*

Data Pasien Rawat Inap 2021 - 2022.csv					
	A	B	C	D	E
1	Nomor Pendaftaran	Jenis Kelamin	Jenis Perawatan	Tanggal Pendaftaran	Ruangan
2	RJ2105030052	PEREMPUAN	Bedah	2021-03-05	Lantai 15
3	RJ2102170209	PEREMPUAN	Bedah	2021-02-17	Lantai 9
4	RJ2101150060	LAKI-LAKI	Bedah	2021-01-15	Lantai 9
5	RJ2205300032	LAKI-LAKI	Bedah Digestif	2022-05-30	B-Lantai 12

Gambar 7. Data Setelah dilakukan *Cleaning*

3.4.2. Data Transformation

Data transformation adalah proses mengubah atau mengkonversi data dari satu bentuk atau format ke bentuk yang lain untuk memudahkan analisis atau pemrosesan [24].

```
# Membuat Labelencoder
label_encoder = LabelEncoder()

# Menetapkan nilai numerik dan menyimpannya di kolom lain yang disebut "Jenis Kelamin_Kode"
df["Jenis Kelamin_Kode"] = label_encoder.fit_transform(df["Jenis Kelamin"])
```

Gambar 8. Potongan Kode *Data Transformation*

```
# Cetak hasil perubahan
print(df['Jenis Kelamin_Kode'])

0      1
1      1
2      0
3      0
4      0
..
19931   0
19932   1
19933   1
19934   0
19935   1
Name: Jenis Kelamin_Kode, Length: 19936, dtype: int64
```

Gambar 9. Data Setelah *Transformation*

3.4.3. Data Selection

Data selection merupakan dari sekumpulan data operasional yang butuh dilakukan saat sebelum sesi penggalian data dalam *Knowledge Discovery in Database* (KDD) diawali Data hasil pilih yang hendak digunakan buat proses data mining, disimpan dalam sesuatu berkas terpisah dari basis data operasional [25].

Data di selection adalah kolom "Jenis Perawatan" sebagai fitur memungkinkan peneliti akan menggunakan informasi di dalamnya sebagai variabel yang akan digunakan dalam pemodelan atau analisis lebih lanjut.

```
features = df['Jenis Perawatan']
features

0      Bedah
1      Bedah
2      Bedah
3  Bedah Digestif
4  Bedah Digestif
...
19931  Urology
19932  Urology
19933  Urology
19934  Urology
19935  Urology
Name: Jenis Perawatan, Length: 19936, dtype: object
```

Gambar 10. Potongan Kode *Data Selection*

3.4.4. Melakukan Pengkodean Variabel Kategorial

Mengencoding variabel kategorikal memungkinkan transformasi nilai kategori menjadi representasi numerik, sehingga dapat digunakan dalam analisis atau pemodelan yang membutuhkan input numerik.

```
label_encoder = LabelEncoder()
features_encoded = label_encoder.fit_transform(features)
```

Gambar 11. Pengkodean Variabel Kategorial

Dengan menggunakan LabelEncoder, variabel kategori dalam features akan diubah menjadi bilangan bulat yang sesuai. Ini diperlukan karena banyak algoritma *machine learning* membutuhkan input numerik, dan dengan mengubah variabel kategori menjadi bilangan bulat, kita dapat menggambarkan informasi kategori dalam bentuk yang dapat diproses oleh algoritma tersebut.

3.4.5. Melakukan Data Normalization

Data normalization merupakan salah satu proses yang dicoba pada fase pra proses data. Pada normalisasi dicoba penskalaan kembali nilai-nilai sehingga bisa membuat pemrosesan lebih gampang [26].

Pada penelitian ini, *data normalization* yang dilakukan menggunakan *Min-Max Scaler* yang merupakan salah satu metode *Scaling* dan memiliki rentang [0, 1].

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

Gambar 12. Melakukan Data Normalization

Fungsi dari kode di atas adalah untuk melakukan penskalaan data pada *dataframe*. Pada objek *MinMaxScaler()* akan dibuat dan disimpan dalam variabel "scaler". *MinMaxScaler* adalah salah satu metode penskalaan yang digunakan untuk mengubah nilai-nilai fitur menjadi rentang yang ditentukan, dalam hal ini, rentang antara 0 dan 1.

Kemudian, array fitur yang sudah di-encode, yaitu "features_encoded", diubah ke bentuk baru menggunakan metode *reshape()*. Metode ini digunakan untuk mengubah bentuk array tanpa mengubah datanya. Dalam hal ini, array "features_encoded" diubah menjadi array 2 dimensi dengan ukuran (-1, 1), yang berarti array akan diubah menjadi satu kolom dan ukuran barisnya disesuaikan secara otomatis sesuai dengan jumlah elemen dalam array. Hasil dari penskalaan fitur disimpan dalam variabel "features_scaled".

3.3. K-Means Clustering

Setelah dilakukan proses pengolahan *data mining*, maka penelitian dimulai dengan melakukan pemodelan terhadap *dataset* data pasien rawat inap.

3.3.1. Menentukan Jumlah Cluster

Pada gambar dibawah ini adalah kodingan yang terkait dengan menentukan nilai k pada algoritma *k-means clustering* salah satunya adalah dengan menghitung wss dan skor silhouette serta membuat plot grafik dari wss dan skor silhouette. Tujuan kedua fungsi ini adalah untuk membantu dalam menentukan jumlah cluster yang paling ideal untuk algoritma *k-means clustering*.

```
# Menentukan Nilai K  
# Membuat fungsi dengan menggunakan skor WSS dan Skor Silhouette untuk nilai k dari 2 hingga kmax  
def calculate_WSS_Silhouette(x, kmax): #points  
    wss=[]  
    silhouetteScore=[]  
    for k in range(2, kmax+1):  
        kmeans = KMeans(n_clusters=k)  
        # Menetapkan Nilai K-Means  
        kmeans.fit(x)  
        # wss  
        wss.append(kmeans.inertia_)  
  
        cluster_labels=kmeans.labels_  
        #Menentukan nilai score silhouette  
        silhouetteScore.append(silhouette_score(x,cluster_labels))  
    return wss, silhouetteScore
```

Gambar 13. Potongan Kode Menentukan Cluster Dengan Menggunakan 2 Metode

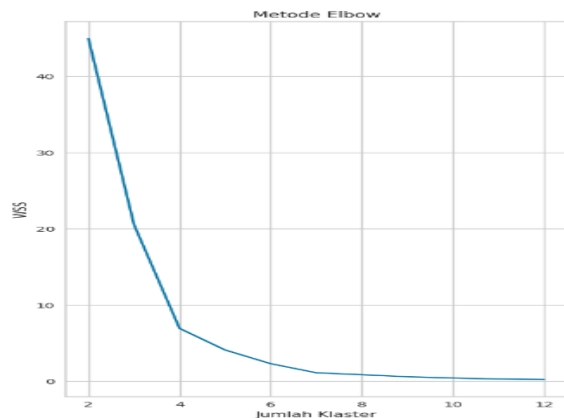
```
# Menentukan Jumlah cluster yang digunakan dengan menggunakan metode WSS dan Silhouette Score  
kmax = 12  
  
def plot_WSS_Silhouette(wss, silhouetteScore, kmax):  
    plt.figure(figsize=(20,10))  
    plt.subplot(1, 3, 3)  
    plt.plot(range(2, kmax+1), wss)  
    plt.title('Metode Elbow')  
    plt.xlabel('Jumlah Klaster')  
    plt.ylabel('WSS')  
    plt.show()  
  
    plt.figure(figsize=(18,10))  
    plt.subplot(1, 2, 2)  
    plt.plot(range(2, kmax+1), silhouetteScore)  
    plt.title('Metode Silhouette')  
    plt.xlabel('Jumlah Klaster')  
    plt.ylabel('Skor Akurasi Metode Silhouette')  
    plt.show()
```

Gambar 14. Potongan Kode Untuk Menampilkan Cluster Dengan Menggunakan 2 Metode

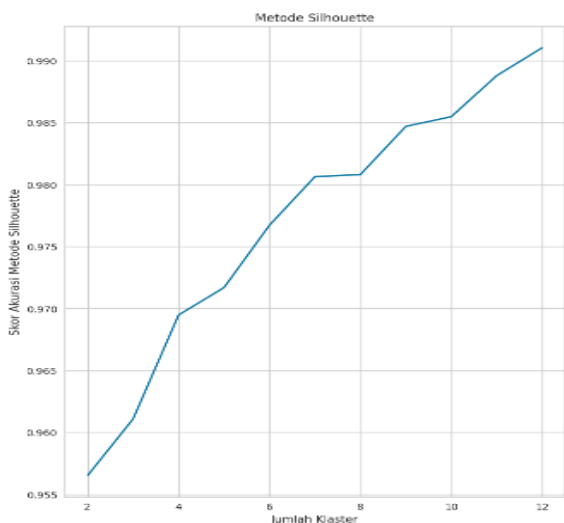
Setelah melakukan proses *cluster* dengan menggunakan *k-means clustering*, selanjutnya adalah evaluasi, dan evaluasi umum digunakan

setelah *k-means clustering* adalah *elbow method* dan *silhouette method* [27], [28].

Grafik yang dihasilkan akan membantu dalam menentukan jumlah *cluster* yang tepat untuk *dataset*. Berikut adalah gambar menampilkan hasil visualisasi dengan menggunakan 2 metode setelah melakukan pemanggilan fungsi tersebut.



Gambar 15. Visualisasi Hasil Evaluasi Jumlah Cluster Metode Elbow



Gambar 16. Visualisasi Hasil Evaluasi Jumlah Cluster Metode Silhouette

Peneliti dapat menyimpulkan bahwa, dengan menggunakan metode wss dan skor silhouette, nilai *cluster* yang memenuhi syarat adalah bentuk lekukan siku pada grafik *cluster*

untuk menerapkan *k-means clustering* dan yang optimal berada pada *cluster* 4.

3.3.2. Membuat Model K-Means Clustering

Pada tahap ini peneliti menentukan jumlah *cluster* dengan baris pertama $k = 4$ menentukan jumlah *cluster* yang diinginkan. Setelahnya peneliti, akan membuat objek KMeans dari pustaka scikit-learn. Objek yang dibuat dengan parameter `n_clusters=k`, yang menentukan jumlah *cluster* yang diinginkan, dan `random_state=42` yang digunakan untuk mengatur inisialisasi acak agar hasilnya dapat direproduksi.

Melakukan fitting pada data `kmeans.fit(features_scaled)` mengaplikasikan algoritma K-Means Clustering pada data yang telah diubah skala (`features_scaled`). Metode `fit()` akan mencari pusat-pusat *cluster* yang optimal dan mengelompokkan setiap sampel ke dalam *cluster* yang sesuai berdasarkan jarak Euclidean atau metrik jarak lainnya.

```
# Memilih jumlah kluster yang akan digunakan
k = 4
# Menginisialisasi untuk pembuatan model K-Means Clustering:
kmeans = KMeans(n_clusters=k, random_state=42)
# Memilih fitur model ke data rekam medis yang telah dipilih:
kmeans.fit(features_scaled)
```

KMeans
KMeans(n_clusters=4, random_state=42)

Gambar 17. Membuat Model Sesuai Jumlah Cluster

Setelah proses fitting selesai, model KMeans akan berisi pusat-pusat *cluster* yang telah ditemukan dan informasi lainnya seperti inersia (SSE, *Sum of Squared Errors*). Model ini dapat digunakan untuk memprediksi *cluster* mana yang sesuai untuk data baru atau untuk melakukan analisis lebih lanjut pada hasil *clustering*.

3.3.3. Membuat Centroid

Pada tahap selanjutnya peneliti akan membuat *centroid*. *Centroid* digunakan untuk mengakses koordinat pusat-pusat *cluster* yang telah ditemukan oleh pada pembuatan model K-Means. Untuk menampilkan hasil *clustering* maka perlu

fungsi `enumerate()`, untuk dapat mengakses indeks dan koordinat pusat dari `cluster_centers`.

```
# Mendapatkan koordinat pusat klaster
cluster_centers = kmeans.cluster_centers_

# 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(cluster_centers):
    print("Pusat Cluster", i, ":", center)
```

Gambar 18. Membuat *Centroid*

Setelah menjalankan program tersebut, maka peneliti dapat lihat hasil dari nilai centroid dan akan dijelaskan pada gambar dibawah ini :

```
Pusat Centroid 0 : [0.75627161]
Pusat Centroid 1 : [0.39423765]
Pusat Centroid 2 : [0.10123967]
Pusat Centroid 3 : [0.96861119]
```

Gambar 19. Hasil Nilai *Centroid*

Dapat dianalisis bahwa dari pusat *centroid* ini adalah bahwa setiap *centroid* mewakili sebuah *cluster* dalam hasil *clustering*. Koordinat nilai *centroid* tersebut menggambarkan pusat atau representasi rata-rata dari sampel-sampel yang termasuk dalam *cluster* tersebut.

3.3.4. Menentukan Jarak Setiap Data Setiap *Centroid*

Setelah melakukan perhitungan *centroid*, tahap selanjutnya peneliti akan menghitung jarak setiap data terhadap setiap *centroid* dan akan ditunjukkan pada gambar potongan kodingan dibawah ini :

```
# Menghitung Jarak Setiap Titik Data

# Data yang akan digunakan
x = np.array([[1, 2], [1, 4], [1, 0], [4, 2], [4, 4], [4, 0]])

# Jumlah klaster yang diinginkan
n_clusters = 4

# Membentuk model K-Means dengan 4 klaster
kmeans = KMeans(n_clusters=n_clusters)
kmeans.fit(features_scaled)

# Mendapatkan pusat klaster
centroids = kmeans.cluster_centers_

# 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 20. Menentukan Jarak Setiap Data Setiap *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 : 1.2673846389382994
Klaster 2 : 1.716222811879347
Klaster 3 : 2.100728665201719
Klaster 4 : 1.0318663322494406
Jarak data 2 ke masing-masing pusat klaster:
Klaster 1 : 3.252872171683422
Klaster 2 : 3.656291855359525
Klaster 3 : 4.00101265270609
Klaster 4 : 3.0315513114843258
```

Gambar 21. Hasil dari Jarak Setiap Data Setiap *Centroid*

Berdasarkan jarak data 1 ke masing-masing pusat cluster dan jarak data 2 ke masing-masing pusat cluster yang diberikan, dapat disimpulkan sebagai berikut:

Jarak Data 1:

- Data 1 memiliki jarak terdekat dengan *cluster* 4, dengan jarak sebesar 1.0318663322494406.
- Data 1 memiliki jarak tertinggi dengan *cluster* 3, dengan jarak sebesar 2.100728665201719.
- Data 1 memiliki jarak menengah dengan *cluster* 1, dengan jarak sebesar 1.2673846389382994.
- Data 1 memiliki jarak sedang dengan *cluster* 2, dengan jarak sebesar 1.716222811879347.

Jarak Data 2:

- Data 2 memiliki jarak terdekat dengan *cluster* 4, dengan jarak sebesar 3.0315513114843258.
- Data 2 memiliki jarak tertinggi dengan *cluster* 3, dengan jarak sebesar 4.00101265270609.
- Data 2 memiliki jarak menengah dengan *cluster* 1, dengan jarak sebesar 3.252872171683422.
- Data 2 memiliki jarak sedang dengan *cluster* 2, dengan jarak sebesar 3.656291855359525.

Dengan demikian, berdasarkan jarak terdekat, kedua data 1 dan data 2 dapat diklasifikasikan sebagai bagian dari *cluster* 4. Sedangkan berdasarkan jarak tertinggi, kedua data tersebut memiliki kemiripan paling rendah dengan *cluster* 3. Dengan demikian data tersebut tidak ada yang berpindah jarak *group*, sehingga penelitian dengan menggunakan algoritma metode *k-means clustering* sudah selesai.

3.3.5. Memvisualisasi Hasil Analisis Dan Pembagian Data Cluster

Pada tahap ini, setelah selesai melakukan proses *cluster* metode *k-means clustering*. Pada tahap ini peneliti akan melakukan menganalisis, memahami, dan menggambarkan pemetaan distribusi data pada setiap *cluster*.

```
# Membuat Visualisasi Hasil Kluster Yang Telah Terbentuk
colors = sns.color_palette("Set3", n_colors=k)

unique_clusters, cluster_counts = np.unique(cluster_labels, return_counts=True)

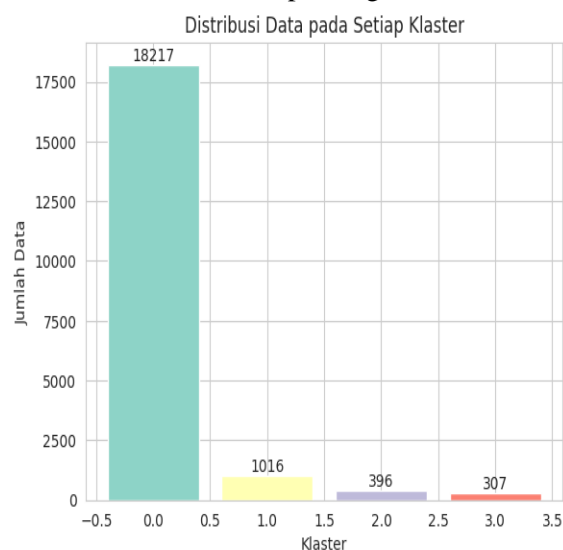
# plt.figure(figsize=(10, 6)) # Comment out or remove this line
plt.bar(unique_clusters, cluster_counts, color=colors)
plt.xlabel('Kluster')
plt.ylabel('Jumlah Data')
plt.title('Distribusi Data pada Setiap Kluster')

# 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 22. Membuat Visualisasi Pembagian Data Cluster

Gambar dibawah ini akan menunjukkan visualisasi hasil analisis pembagian data *cluster*:



Gambar 23. Hasil Visualisasi Pembagian Data Setiap Cluster

Pada gambar diatas menunjukkan hasil pembagian data setiap *cluster* bahwa 18217 (91,4%) data pasien dari 19936 total data pasien yang berada di *cluster* 1, 1016 (5.1%) data pasien dari 19936 total data pasien berada di *cluster* 2, 396 (2%) data pasien dari 19936 total data pasien berada di *cluster* 3, dan 307 (1,5%) data pasien dari 19936 total data pasien berada di *cluster* 4.

3.3.6. Hasil Analisa Data Setiap Cluster

Hasil akhir ditampilkan dalam tabel berikut berdasarkan data yang ditunjukkan pada tabel dibawah :

Tabel 1. Hasil Analisa Data Pada Cluster 1 Dan Cluster 2

Hasil Analisa Cluster 1	Hasil Analisa Cluster 2
Terdiri dari 18217 Orang yang berasal dari, Jenis Perawatan : 1. Pelayanan Gawat Darurat : 17796 2. Orthopedi : 280 3. Obgyn : 64 4. Penyakit Dalam : 55 5. Paru – Paru : 14 6. NICU / PISU : 2 7. Penyakit Jantung : 2 8. Perinatologi : 2 9. Neonatal & Tumbuh Kembang : 1 10. Onkologi Kebidanan : 1	Terdiri dari 1016 Orang yang berasal dari, Jenis Perawatan : 1. Gigi & Mulut : 413 2. Kemoterapi : 368 3. Kesehatan Anak : 221 4. Mata : 6 5. Gastro Entero Hepatologi : 2 6. Isolasi : 2 7. Endoscopy : 1 8. ICU : 1 9. Kesehatan Syaraf : 1 10. Kulit & Kelamin : 1
Jenis Kelamin : 1. Perempuan : 9227 2. Laki – Laki : 8990	Jenis Kelamin : 1. Perempuan : 583 2. Laki – Laki : 433

Tabel 2. Hasil Analisa Data Pada Cluster 3 Dan Cluster 4

Hasil Analisa Cluster 3	Hasil Analisa Cluster 4
Terdiri dari 396 Orang yang berasal dari, Jenis Perawatan : 1. Bedah Onkologi : 172 2. Bedah Digestif : 58 3. Bedah Saraf : 57 4. Bedah Umum : 48 5. Bedah Plastik : 28 6. Bedah Syaraf : 21 7. Endokrin : 7 8. Bedah : 3 9. Bedah Orthopedi : 2	Terdiri dari 417 Orang yang berasal dari, Jenis Perawatan : 1. Urology : 147 2. THT : 129 3. Umum : 19 4. Saraf : 7 5. Reumatologi : 5
Jenis Kelamin : 1. Perempuan : 228 2. Laki – Laki : 168	Jenis Kelamin : 1. Perempuan : 112 2. Laki – Laki : 195

Dari tabel di atas, kita dapat menemukan empat kolom cluster. Cluster 1 terdiri dari 1827 orang yang menderita penyakit Pelayanan Gawat Darurat, Orthopedi, Obgyn, Penyakit Dalam, Paru-Paru, NICU/PISU, Penyakit Jantung, Perinatologi, Neonatal dan Tumbuh Kembang, Onkologi Kebidanan, dan laki-laki dan perempuan masing-masing 9227 dan 8990. Cluster 4 terdiri dari 417 orang yang menderita penyakit Urologi, THT, Umum, Saraf, Reumatologi, dan jenis kelamin laki-laki sebanyak 195 orang dan perempuan sebanyak 112 orang.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, penerapan *data mining* untuk Pengelolaan data rekam medis menggunakan metode *k-means clustering* pada rumah sakit royal prima medan, maka dapat diambil kesimpulannya sebagai berikut, Peneliti dapat menyimpulkan hasil analisis bahwa data rekam medis dapat dikelompokkan menjadi 4 *cluster* dengan 2 variabel, yaitu : jenis kelamin dan jenis perawatan dan hasil analisis ini memperoleh hasil *cluster* 1 sebanyak 18217 pasien, *cluster* 2 sebanyak 1016 pasien, *cluster* 3 sebanyak 396 pasien dan *cluster* 4 sebanyak 307 pasien yang di rawat inap pada Rumah Sakit Royal Prima Medan.

5. REFERENSI

- [1] C. Sanggamele, F. K. Kolibu, and F. R. R. Maramis, "ANALISIS PENGELOLAAN REKAM MEDIS DI RUMAH SAKIT UMUM PANCARAN KASIH MANADO."
- [2] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit Anwar Medika Balong Bendo Sidoarjo," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 19, no. 1, pp. 186–195, 2019, doi: 10.30812/matrik.v19i1.529.
- [3] I. Oktanisa and A. A. Supianto, "Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, p. 567, 2018, doi: 10.25126/jtiik.201855958.
- [4] P. Kasih, "Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara," *Innov. Res. Informatics*, vol. 1, no. 2, pp. 63–69, 2019, doi: 10.37058/innovatics.v1i2.918.
- [5] D. S. O. Panggabean, E. Buulolo, and N. Silalahi, "Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda," *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 56, Feb. 2020, doi: 10.30865/jurikom.v7i1.1947.
- [6] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>
- [7] J. Wandana, S. Defit, and S. Sumijan, "Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means," *J. Inf. dan Teknol.*, vol. 2, pp. 4–9, 2020, doi: 10.37034/jidt.v2i4.73.
- [8] J. P. Gultom and A. Rikki, "Implementasi Data Mining menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin," *Kumpul. Artik. Karya Ilm. Fak. Ilmu Komput.*, vol. 02, no. 01, 2020.
- [9] W. Purba, W. Siawin, and . H., "Implementasi Data Mining Untuk Pengelompokan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 85–90, 2019, doi: 10.34012/jusikom.v2i2.429.
- [10] I. T. Julianto, D. Kurniadi, M. R. Nashrulloh, and A. Mulyani,

- "Comparison of Data Mining Algorithm For Forecasting Bitcoin Crypto Currency Trends," *J. Tek. Inform.*, vol. 3, no. 2, pp. 245–248, 2022, doi: 10.20884/1.jutif.2022.3.2.194.
- [11] S. P. Tamba, M. Sitanggang, B. C. Situmorang, and G. Laura, "APPLICATION OF DATA MINING TO DETERMINE THE LEVEL OF FISH SALES IN PT . TRANS RETAIL WITH FP-GROWTH METHOD," vol. 10, no. 2, pp. 905–913, 2022.
- [12] A. Riani, Y. Susianto, and N. Rahman, "Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes," *J. Innov. Inf. Technol. Appl.*, vol. 1, no. 01, pp. 25–34, 2019, doi: 10.35970/jinita.v1i01.64.
- [13] W. Purba, "Penerapan Data Mining Menggunakan Metode Apriori Untuk Menyusun Pola Persediaan Inventaris Barang Pada Biro Sarana Dan Prasarana Universitas Prima Indonesia," *J. Sekol. POSD FIP UNIMED*, vol. 3, no. 9, pp. 191–196, 2019.
- [14] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [15] M. Sari, "Penelitian Kepustakaan (Library Research) dalam Penelitian Pendidikan IPA."
- [16] 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, Oct. 2020, doi: 10.33060/jik/2020/vol9.iss2.181.
- [17] A. Yati, "METODE KOMUNIKASI DA'I PERBATASAN ACEH SINGKILDALAM MENJAWABTANTANGAN DAKWAH".
- [18] W. N. Purba, D. Situmorang, Y. Alfani, D. Hutabarat, and F. W. Anggiono, "Implementasi Data Mining Dengan Metode Pohon Keputusan Algoritma Id3 Untuk Menentukan Memprediksi Penjualan Pada CV. Mitra Baja Cemerlang," *J. Mantik Penusa*, vol. 2, no. 1, pp. 82–86, 2019.
- [19] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri)," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 5, no. 2, pp. 80–85, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2393.
- [20] S. P. Tamba, A. W. Tan, Y. Gunawan, and ..., "Penerapan Data Mining Untuk Pembuatan Paket Promosi Penjualan Menggunakan Kombinasi Fp-Tree Dan Tid-List," ... (*Teknik Inf. dan ...*, vol. 4, 2021.
- [21] B. Budiman, "Perbandingan Algoritma Klasifikasi Data Mining untuk Penelusuran Minat Calon Mahasiswa Baru," *Nuansa Inform.*, vol. 15, no. 2, pp. 37–52, 2021, doi: 10.25134/nuansa.v15i2.4162.
- [22] N. L. P. P. Dewi, I. N. Purnama, and N. W. Utami, "Penerapan Data Mining Untuk Clustering Penilaian Kinerja Dosen Menggunakan Algoritma K-Means (Studi Kasus: STMIK Primakara)," *J. Ilm. Teknol. Inf. Asia*, vol. 16, no. 2, p. 105, Jul. 2022, doi: 10.32815/jitika.v16i2.761.
- [23] H. Erasmus Simanjuntak, "JURNAL MEDIA INFORMATIKA BUDIDARMA Analisa Data Mining Menggunakan Frequent Pattern Growth pada Data Transaksi Penjualan PT Mora Telematika Indonesia untuk Rekomendasi Strategi Pemasaran Produk Internet", doi:

- 10.30865/mib.v4i4.2300.
- [24] M. B. Abdul Majid, Y. M. Cani, and U. Enri, "Penerapan Algoritma K-Means dan Decision Tree Dalam Analisis Prestasi Siswa Sekolah Menengah Kejuruan," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 355, Dec. 2022, doi: 10.30865/json.v4i2.5299.
- [25] P. Bidang Komputer Sains dan Pendidikan Informatika, D. Akademi Perekam dan Informasi Kesehatan Iris Padang Jl Gajah Mada No, and S. Barat, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi".
- [26] J. Homepage, I. Permana, and F. Nur Salisah, "IJIRSE: Indonesian Journal of Informatic Research and Software Engineering The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation".
- [27] D. Ayu, I. C. Dewi, and K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," 2019.