ADIJAYA
Jurnal Multidisiplin

https://e-journal.naureendigition.com/index.php/mj

Vol. 0 4 , No. 02, 2026, Hal. 134-144

E-ISSN : 2964-7606

KNOWLEDGE PROCESSING MENGGUNAKAN CLUSTERING PADA IDENTIFIKASI KETERSEDIAAN FASILITAS KESEHATAN TERHADAP KELUHAN MASYARAKAT INDONESIA TAHUN 2025

Bismi Padia[1], Farhan Hamdallah[2]
Email: bismiipadiaa@gmail.com
Sistem Informasi, STMIK Likmi, Kota Bandung.
Sistem Informasi, Universitas Sains Indonesia, Kota Bekasi.

Abstrak

Informasi adalah data yang telah diolah dengan cara tertentu dan memiliki nilai yang relevan secara kontekstual untuk digunakan dalam pengambilan keputusan, sedangkan data adalah fakta mentah yang tidak memiliki makna. Hubungan antara pengetahuan, data, dan informasi menunjukkan bahwa pengetahuan akan terus berkembang selama proses pengambilan keputusan. Studi ini menganalisis ketersediaan fasilitas kesehatan terhadap keluhan masyarakat di Indonesia pada tahun 2025. Data yang digunakan meliputi jumlah tenaga medis, keluhan masyarakat, dan fasilitas kesehatan di berbagai wilayah. Algoritma K-Means digunakan untuk mengelompokkan data studi berdasarkan atribut yang sebanding. Skor silhouette juga digunakan untuk menghitung jumlah cluster yang ideal. Hasil penelitian menunjukkan bahwa penemuan pola tertentu memungkinkan transformasi data menjadi pengetahuan. Data yang telah dikumpulkan dimasukkan ke dalam beberapa kategori yang menunjukkan ketersediaan fasilitas kesehatan serta tingkat keluhan masyarakat. Terdapat tiga kategori: rendah, sedang, dan tinggi. Akibatnya, pengolahan pengetahuan memiliki kemampuan untuk memberikan pemahaman yang lebih mendalam tentang kondisi lapangan yang sebenarnya. Selain itu, proses pengolahan pengetahuan dalam penelitian ini melibatkan interpretasi pola dan pengambilan makna dari hasil clustering; ini menunjukkan tingkat kompleksitas yang lebih tinggi dibandingkan dengan pengolahan data dan informasi biasa. Penggunaan teknik pengolahan data dan pengelolaan database merupakan komponen penting dalam mendukung proses pengolahan pengetahuan, sehingga hasil yang diperoleh dapat digunakan sebagai dasar dalam pengambilan keputusan terkait.

Kata Kunci: Ketersediaan, Fasilitas Kesehatan, Teknologi .

Abstract

Information is data that has been processed in a certain way and has contextually relevant value for use in decision making, while data is raw facts that do not have meaning. The relationship between knowledge, data, and information shows that knowledge will continue to develop during the decision-making process. This study analyzes the availability of healthcare facilities in relation to public complaints in Indonesia in 2025. The data used include the number of medical personnel, public complaints, and healthcare facilities in various regions. The K-Means algorithm is used to group the study data based on comparable attributes. The silhouette score is also used to determine the ideal number of clusters. The results show that identifying certain patterns enables the transformation of data into knowledge. The collected data are grouped into several categories that indicate the availability of healthcare facilities as well as the level of public complaints. There are three categories: low, medium, and high. As a result, knowledge processing has the ability to provide a deeper understanding of actual field conditions. In addition, the knowledge processing in this study involves pattern interpretation and deriving meaning from clustering results; this indicates a higher level of complexity compared to ordinary data and information processing. The use of data processing techniques and database management is an important component in supporting the knowledge processing, so that the results obtained can be used as a basis for related decision making.

***Keywords**:* Availability, Healthcare Facilities, Technology

Pendahuluan

Sejak lama, pemrosesan pengetahuan telah memainkan peran penting dalam kemajuan ilmu pengetahuan dan teknologi, khususnya dalam membantu pengambilan keputusan berbasis data. Pada era modern, pemrosesan pengetahuan berkembang pesat karena kemajuan dalam pengelolaan basis data, penggalian data, dan pembelajaran mesin. Gagasan ini telah mengilhami berbagai disiplin ilmu, termasuk pengenalan pola, analisis data, sistem pendukung keputusan, dan kecerdasan buatan (AI). Disiplin ilmu ini menawarkan kerangka kerja yang memungkinkan para peneliti dan praktisi mengolah sejumlah besar data menjadi pengetahuan dan informasi yang bermanfaat. Untuk menentukan kualitas layanan yang diberikan kepada masyarakat, ketersediaan tenaga medis dan fasilitas kesehatan sangat penting, terutama di sektor kesehatan. Pemrosesan pengetahuan dapat memberikan gambaran yang lebih baik tentang situasi di berbagai wilayah dengan mengolah data tentang fasilitas kesehatan, tenaga medis, dan keluhan masyarakat. Ketidakseimbangan dalam jumlah fasilitas kesehatan, tenaga medis, dan keluhan masyarakat menunjukkan ketidakseimbangan dalam pemerataan pelayanan kesehatan. Clustering, teknik data mining, dapat digunakan untuk mengelompokkan data berdasarkan karakteristik yang sebanding. Data dapat dikelompokkan ke dalam beberapa cluster dengan

algoritma K-Means untuk menunjukkan kondisi tertentu, seperti tingkat keluhan masyarakat dan ketersediaan fasilitas kesehatan. Melalui proses ini, pola-pola tersembunyi yang sebelumnya sulit diketahui secara langsung dapat diidentifikasi. Selain itu, teknik evaluasi seperti skor silhouette membantu dalam menentukan jumlah cluster yang ideal; hasil pengelompokan yang dihasilkannya lebih akurat dan representatif. Pengelompokan data kesehatan ini serupa dengan proses pengenalan pola, di mana sistem dapat mengidentifikasi area berdasarkan ciri-cirinya. Namun, hasil pengelompokan tidak selalu bersifat pasti karena variasi data dan kemungkinan ketidakpastian. Akibatnya, metode matematis dan statistik digunakan untuk meningkatkan kepercayaan terhadap hasil analisis. Akibatnya, penelitian ini tidak hanya memproses data tetapi juga menginterpretasikan temuan untuk meningkatkan pemahaman. Hasilnya diharapkan dapat membantu pengambilan keputusan, khususnya tentang perencanaan dan pemerataan fasilitas kesehatan di Indonesia, dengan mempertimbangkan data aktual.

Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan analisis data menggunakan pembelajaran mesin, khususnya metode clustering. Karena penelitian ini berfokus pada pengolahan data numerik dan bertujuan untuk menghasilkan pengelompokan yang objektif yang didasarkan pada perhitungan matematis, pendekatan kuantitatif dipilih. Penelitian deskriptif kuantitatif menggunakan data yang diperoleh untuk menemukan pola dan ciri-ciri tertentu tanpa mengubah variabel yang ada. Clustering, metode pembelajaran tidak diawasi, digunakan selama analisis. Data dikelompokkan berdasarkan tingkat kemiripan antar data. Penelitian ini menggunakan data sekunder open source dari Badan Pusat Statistik (BPS). Tiga variabel utama termasuk jumlah tenaga kesehatan, fasilitas kesehatan, dan keluhan masyarakat tentang kesehatan. Setelah itu, data diproses melalui tahapan pra-pemrosesan, yang mencakup pembersihan, normalisasi, dan transformasi, sehingga dapat digunakan untuk analisis dengan menggunakan metode clustering. Selanjutnya, algoritma clustering (seperti K-Means) digunakan untuk melakukan proses pengelompokan data. Algoritma ini mengelompokkan data ke dalam berbagai cluster berdasarkan jarak terdekat antar data. Dalam penelitian ini, empat kelompok dibentuk: rendah, sedang, tinggi, dan sangat tinggi. Setiap data akan dimasukkan ke dalam kumpulan yang memiliki atribut yang paling dekat satu sama lain. Metode evaluasi seperti Silhouette Score digunakan untuk mengukur seberapa baik data dikelompokkan dalam cluster yang tepat. Hasil clustering memiliki tingkat akurasi dan pemisahan yang baik, seperti yang ditunjukkan oleh nilai evaluasi ini. Metode penelitian ini diharapkan dapat menghasilkan pengelompokan data representatif pada tahap ini. Ini juga dapat digunakan sebagai dasar untuk analisis lebih lanjut terkait kondisi kesehatan masyarakat dengan data yang saat ini tersedia.

Hasil dan Pembahasan

a. Data Preprocessing

Data diambil dari website Badan Pusat Statistik sehingga didapat data yang belum dilakukan preprocessing. Data diambil secara langsung1 dan masih mentah sehingga belum dilakukan preprocessing mulai dari missing value, data duplikat, dan lain-lain.

1. Import Library

```
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
import math
import plotly as py
import plotly.graph_objs as go
import warnings
import os
import joblib
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from yellowbrick.cluster import KElbowVisualizer
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from google.colab import files
from sklearn.decomposition import PCA
warnings.filterwarnings("ignore")
py.offline.init_notebook_mode(connected = True)
```

Gambar 1. Import Library

Import Library diperlukan untuk menjalankan program, algoritma machine learning yang diperlukan, tools visualisasi menggambarkan data, dan lainnya. Sehingga dengan adanya Import Library akan mempermudah kita melakukan pemrosesan data mulai dari preprocessing hingga output akhir yang diperlukan.

2. Mengetahui Data

```
df = pd.read_csv(url)
df.head()
```

	Provinsi	Total_Nakes	Total_RS	Total_Puskesmas	Keluhan
0	Aceh	623	86.0	366.0	2460
1	Sumatera Utara	1634	208.0	619.0	2432
2	Sumatera Barat	1090	83.0	280.0	2847
3	Riau	1823	81.0	242.0	2174
4	Jambi	2131	45.0	208.0	2595

Tabel 1. Data Frame Head

Sebelum melakukan pengolahan data lainnya, lebih baik kita mengetahui isi data yang telah kita ambil dari Badan Pusat Statistik, dimana pada gambar hanya ingin mengetahui 5 baris pada data paling atas.

3. Info Nilai dari Setiap Data

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39 entries, 0 to 38
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Provinsi         39 non-null     object
 1   Total_Nakes      39 non-null     int64
 2   Total_RS         39 non-null     float64
 3   Total_Puskesmas  39 non-null     float64
 4   Keluhan          39 non-null     int64
dtypes: float64(2), int64(2), object(1)
memory usage: 1.7+ KB
```

Gambar 2. Data Frame Info

Dari sebuah data ini kita memerlukan suatu informasi yang bisa membantu mulai dari tipe data seperti integer, float, char, dan lainnya. Hal ini mempermudah untuk melakukan pemrosesan data seperti integer dan float kita bisa melakukan pengolahan data secara statistik karena mengandung angka-angka yang bisa dilakukan perhitungan.

4. Mengetahui Statistika dari Data

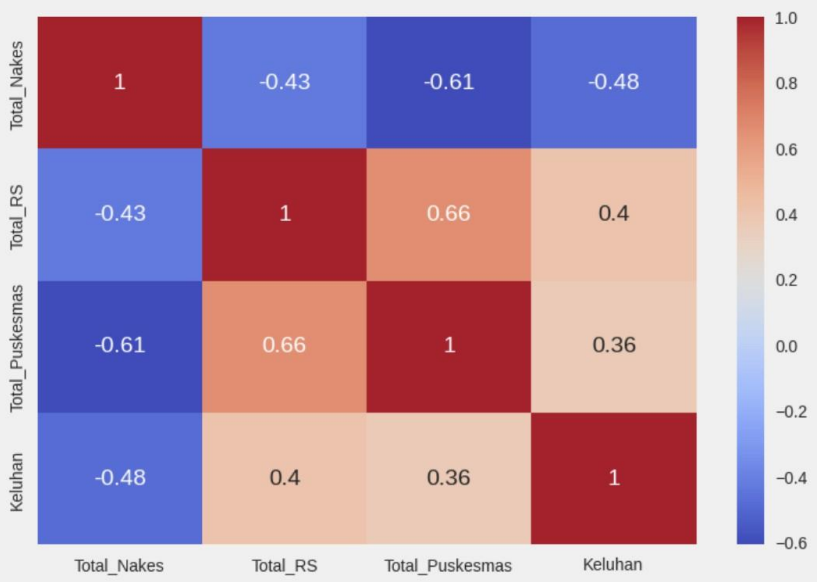
```
df.describe(include="all")
```

	Provinsi	Total_Nakes	Total_RS	Total_Puskesmas	Keluhan
count	39	39.000000	39.000000	39.000000	39.000000
unique	39	NaN	NaN	NaN	NaN
top	Aceh	NaN	NaN	NaN	NaN
freq	1	NaN	NaN	NaN	NaN
mean	NaN	1925.846154	97.455641	264.366667	2433.128205
std	NaN	969.858360	127.304521	247.210141	659.208224
min	NaN	246.000000	8.000000	10.300000	857.000000
25%	NaN	1050.000000	25.500000	119.500000	2024.500000
50%	NaN	2007.000000	52.000000	203.000000	2460.000000
75%	NaN	2393.000000	84.500000	294.000000	2908.500000
max	NaN	4122.000000	516.770000	1106.000000	4200.000000

Tabel 2. Data Frame Describe

Data Frame Describe untuk mengetahui hitungan statistika secara angka mulai dari data paling besar, data paling kecil, rata-rata, dan lainnya seperti pada gambar. Informasi ini diperlukan untuk mengetahui data sudah siap atau belum untuk dilakukan pengolahan secara matematis.

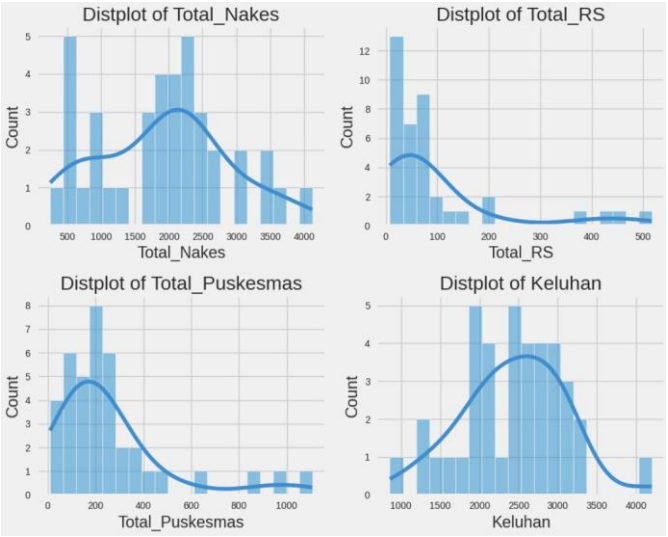
5. Matriks Korelasi



Gambar 3. Matriks Korelasi

Matriks Korelasi diperlukan untuk mengetahui identifikasi hubungan antar variabel lainnya. Seperti pada gambar 5 antara Total_RS dengan Total_Puskesmas memiliki hubungan yang terbilang kuat yaitu sekitar 0,66 dan bernilai positif sehingga miliki hubungan antar variabelnya. Nilai terbesar kedua ditempati oleh variabel Total_RS dengan variabel Keluhan yang memiliki nilai 0,4 dan terbilang ada hubungan namun lemah dan tidak sekuat hubungan variabel sebelumnya yang dibahas.

6. Diagram Batang



Gambar 4. Diagram Batang Masing-Masing Variabel Numerik

Dengan dibentuknya diagram batang setiap variabel yang bernilai numerik menjadikan data lebih mudah dilihat dari kecondongannya (skewness), hal ini untuk mengidentifikasi nilai modus, median, dan mean setiap variabelnya. Pada data semua variabel termasuk skewness positif yang artinya nilai Mean > Median > Modus. Namun jika diperhatikan dengan benar, data yang ada pada variabel Keluhan seperti mendekati Data Normal.

7. Missing Values

```
missing_values = df.isnull().sum()
missing_values[missing_values > 0]
```

```
Series([], dtype: int64)
```

Gambar 5. Missing Value

Missing Value untuk mengetahui apakah data pada masing baris dan kolomnya terdapat data yang hilang atau tidak, jika ada data yang hilang maka akan dilakukan pengisian dengan metode Imputasi Median atau Modus. Namun karena di data ini tidak ada Missing Value maka Metode Imputasi, Dropna, atau Metode lainnya tidak dilakukan.

8. Duplicated

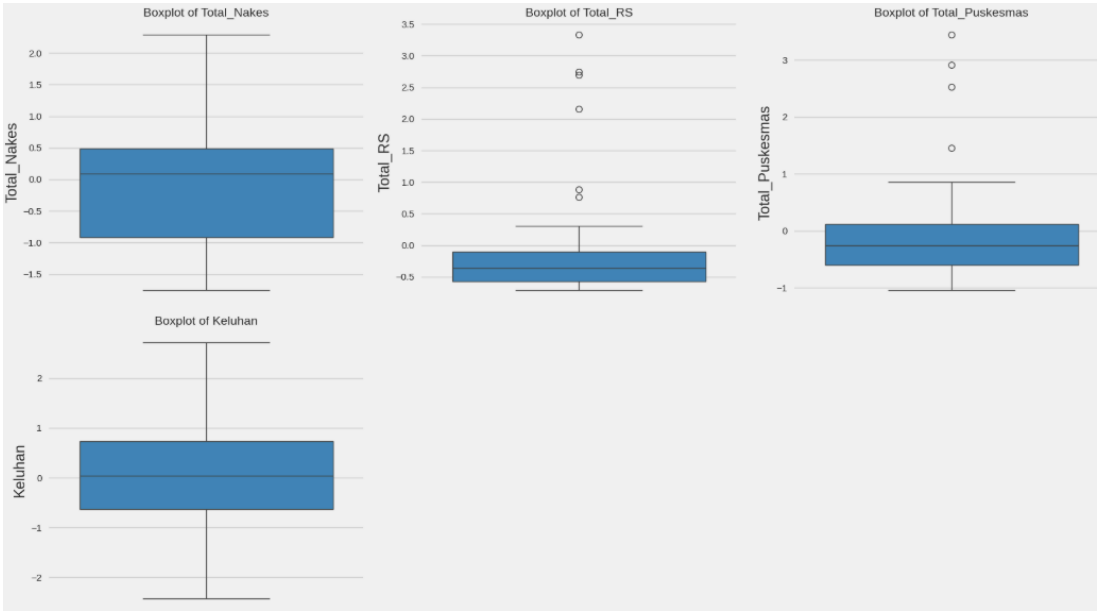
```
duplicated = df.duplicated().sum()

print("Jumlah data duplikat:", duplicated)
```

```
Jumlah data duplikat: 0
```

Gambar 6. Data Frame Duplicated

Data Frame Duplicated bertujuan mengetahui data yang sama secara baris di setiap kolomnya, namun setelah dicek dari data yang ada tidak ditemukan data duplikasi atau double sehingga tidak diperlukan langkah selanjutnya untuk melakukan penghapusan pada data yang memiliki duplikasi tersebut.

9. Identifikasi Outlier



Gambar 7. Boxplot Tiap Variabel Numerik

Identifikasi Outlier menggunakan Boxplot digunakan pada data ini dikarenakan jumlah data yang terbilang banyak dari seluruh Indonesia, maka diperlukan untuk mengidentifikasi data yang berbeda sendiri atau Outlier. Pada Gambar 7 terdapat Outlier yang terdapat pada variabel Total_RS dan Total_Puskesmas. Data Outlier ini tidak akan mempengaruhi pada langkah selanjutnya karena data pada penelitian ini akan diambil dari data yang sudah ada di Badan Pusat Statistik agar hasilnya juga lebih realistis sesuai dengan kondisi pada masyarakat.

b. Label Encoder

Total_Nakes	NakesRange	Nakes_encoded	Total_RS	RSRange	RS_encoded	Total_Puskesmas	PuskesmasRange	Puskesmas_encoded	Keluhan
-1.360897	Rendah	0	-0.091162	Sangat Tinggi	1	0.416496	Sangat Tinggi	1	0.041297
-0.304850	Sedang	2	0.879697	Sangat Tinggi	1	1.453295	Sangat Tinggi	1	-0.001734
-0.873089	Sedang	2	-0.115036	Tinggi	3	0.064066	Tinggi	3	0.636039
-0.107429	Sedang	2	-0.130952	Tinggi	3	-0.091659	Tinggi	3	-0.398229
0.214295	Tinggi	3	-0.417435	Sedang	2	-0.230992	Tinggi	3	0.248765
-1.469531	Rendah	0	-0.067289	Sangat Tinggi	1	0.367319	Sangat Tinggi	1	0.282575
0.412761	Tinggi	3	-0.552719	Sedang	2	-0.349834	Sedang	2	0.863486
0.084770	Sedang	2	-0.115036	Tinggi	3	0.236183	Sangat Tinggi	1	0.866559
0.088948	Tinggi	3	-0.544761	Sedang	2	-0.821107	Rendah	0	1.092469
0.764777	Sangat Tinggi	1	-0.465182	Sedang	2	-0.689970	Rendah	0	-0.992971

Gambar 8. Label Encoder

Encoding dilakukan pada pengolahan data ini agar lebih sepadan dan tidak ada perbedaan jumlah data yang signifikan antara variabel satu dengan yang lainnya. Sebelum dilakukan Label Encoder dilakukan pembagian Binning terlebih dahulu menggunakan otomatis metode Q-Cut sehingga terbagi masing-masing variabel menjadi 4 kelompok yaitu Rendah, Sedang, Tinggi, dan Sangat Tinggi.

c. Unsupervised Learning Menggunakan Clustering

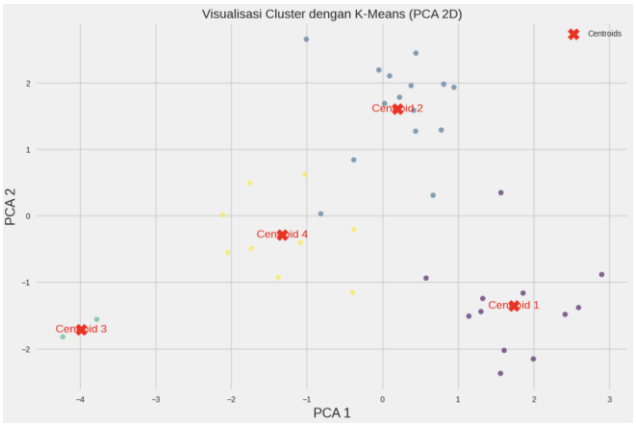
Unsupervised Learning terdapat pada Data Minning dan juga termasuk pada Machine Learning dimana data yang tadinya hanya memiliki nilai saja, lalu dilakukan pengolahan data, identifikasi pola dari data, lalu kita bisa mengambil kesimpulan data termasuk kedalam kelompok mana dan akan dibagi menjadi berapa kelompok, sehingga masing-masing dari variabel akan memiliki pembagian nilai sesuai dengan rangenya dan memiliki label sesuai dengan kelompoknya.

d. K-Means

Gambar 9. Algoritma K-Means

K-Means adalah salah satu Algoritma yang ada pada Unsupervised Learning dimana Algoritma ini memiliki keunggulan yaitu bersifat *non-hierarchical*, cepat, dan mempartisi data dengan mencari titik pusat (centroid) terbaik secara iteratif untuk meminimalkan jarak antara titik data dan centroid-nya. Hasil dari Algoritma ini adalah mengelompokkan data menjadi 4 cluster.

e. Silhouette Score



Gambar 10. Silhouette Score

Silhouette Score diperlukan untuk mengukur seberapa baik pembagian cluster pada Algoritma K-Means, dan didapat hasil yang sama yaitu jumlah Centroid sebanyak 4.

	Total_Nakes	Total_RS	Total_Puskesmas	Keluhan	Target
0	-1.360897	-0.091162	0.416496	0.041297	4
1	-0.304850	0.879697	1.453295	-0.001734	4
2	-0.873089	-0.115036	0.064066	0.636039	2
3	-0.107429	-0.130952	-0.091659	-0.398229	2
4	0.214295	-0.417435	-0.230992	0.248765	2
5	-1.469531	-0.067289	0.367319	0.282575	4
6	0.412761	-0.552719	-0.349834	0.863486	2
7	0.084770	-0.115036	0.236183	0.866559	2
8	0.088948	-0.544761	-0.821107	1.092469	2
9	0.764777	-0.465182	-0.689970	-0.992971	1
10	-0.605682	0.768287	-0.903067	0.151946	4

Tabel 3. Penerapan Silhouette Score Pada Data

Dari Gambar 11 terlihat bahwa pada data sudah terdapat identifikasi sesuai dengan cluster yang dibagi pada Algoritma sebelumnya yaitu dengan nama kolom Target.

f. Data Frame Invers

	Provinsi	Total_Nakes	Total_RS	Total_Puskesmas	Keluhan	NakesRange	RSRange	PuskesmasRange	KeluhanRange
0	Aceh	623.0	86.0	366.0	2460.0	Rendah	Sangat Tinggi	Sangat Tinggi	Sedang
1	Sumatera Utara	1634.0	208.0	619.0	2432.0	Sedang	Sangat Tinggi	Sangat Tinggi	Sedang
2	Sumatera Barat	1090.0	83.0	280.0	2847.0	Sedang	Tinggi	Tinggi	Tinggi
3	Riau	1823.0	81.0	242.0	2174.0	Sedang	Tinggi	Tinggi	Sedang
4	Jambi	2131.0	45.0	208.0	2595.0	Tinggi	Sedang	Tinggi	Tinggi

Tabel 4. Data Frame Invers

Data Frame Invers bertujuan untuk mengembalikan pada nilai yang asli sesudah data diolah dan dikelompokkan sesuai dengan clusternya. Misalkan terlihat bahwa Total_Nakes pada Aceh sebanyak 623 termasuk ke dalam NakesRange Rendah, hal ini berlaku untuk nilai dari variabel lainnya.

Kesimpulan

Studi ini menunjukkan bahwa teknik clustering, terutama algoritma K-Means, dapat digunakan untuk pemrosesan pengetahuan (knowledge processing) untuk mengubah data mentah tentang tenaga medis, keluhan masyarakat, dan fasilitas kesehatan menjadi informasi yang lebih bermakna. Dalam proses ini, data yang dihasilkan dikelompokkan ke dalam kategori rendah, sedang, tinggi, dan sangat tinggi. Kategori ini menunjukkan kondisi

ketersediaan layanan kesehatan yang sebenarnya di berbagai wilayah di Indonesia. Pembagian menjadi empat cluster sudah cukup optimal dan mampu menemukan pola dan hubungan antar variabel, menurut hasil clustering yang divalidasi menggunakan Silhouette Score. Oleh karena itu, penelitian ini tidak hanya melakukan pengolahan data; itu juga menemukan pola ketimpangan antara tingkat keluhan masyarakat dan ketersediaan fasilitas kesehatan. Secara keseluruhan, penelitian ini menunjukkan bahwa machine learning dan data mining dalam pengolahan pengetahuan sangat membantu dalam pengambilan keputusan. Hasil yang diperoleh dapat digunakan sebagai dasar untuk perencanaan dan pemerataan fasilitas kesehatan di Indonesia. Ini akan membuat kebijakan yang dibuat lebih tepat sasaran dan sesuai dengan keadaan lapangan yang sebenarnya.

Daftar Referensi

Badan Pusat Statistik. (2025). *Data fasilitas kesehatan dan tenaga medis di Indonesia*.

Jiawei Han, Micheline Kamber, & Jian Pei. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Trevor Hastie, Robert Tibshirani, & Jerome Friedman. (2009). *The Elements of Statistical Learning*. Springer.

Christopher M. Bishop. (2006). *Pattern Recognition and Machine Learning*. Springer.

Ian H. Witten, Eibe Frank, & Mark A. Hall. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

Tom M. Mitchell. (1997). *Machine Learning*. McGraw-Hill.

Gareth James, Daniela Witten, Trevor Hastie, & Robert Tibshirani. (2013). *An Introduction to Statistical Learning*. Springer.

Pang-Ning Tan, Michael Steinbach, & Vipin Kumar. (2006). *Introduction to Data Mining*. Addison-Wesley.

Kementerian Kesehatan Republik Indonesia. (2024). *Profil Kesehatan Indonesia*.