

PENERAPAN ALGORITMA KMEANS CLUSTERING – CALINSKI HARABASZ DALAM MENDIAGNOSA PENYAKIT HIPERTENSI

(The Implementation of K-means Clustering Algorithm - Calinski Harabasz in Diagnosing Hypertension Disease)

Korisaputri^{1,*}, Khoirun Nisa², Purwono³

Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Harapan Bangsa,
Jl. K.H. Wahid Hasyim, No. 274, Windusara, Karangklesem, Kec. Purwokerto Selatan, Kab.Banyumas, Jawa Tengah
53144, Indonesia.

¹chorisaputri@gmail.com*; ²khairunnisa@uhb.ac.id; ³purwono@uhb.ac.id

ABSTRACT

The increasing number and variety of diseases affecting society due to changes in lifestyle influenced by the progress of the times have become a major factor in causing deaths worldwide. One of the Non-Communicable Diseases (NCDs) that has now become a serious challenge in the field of health is hypertension. Hypertension, also known as high blood pressure, refers to a condition in which the pressure in the arteries increases. This research focuses on clustering and diagnosing hypertension and aims to perform clustering on the data. The methodology used in this research is based on CRISP-DM (Cross-Industry Standard Process for Data Mining). The results of the experiments are then evaluated using Calinski-Harabasz (CH) metrics. The evaluation value produced by Calinski-Harabasz (CH) is 371.70 for cluster (k)=3. The results indicate that the K-Means algorithm produces optimal cluster grouping quality, as evidenced by the high CH value. The higher the CH value, the better the cluster grouping.

Keywords : Clustering, K-Means, Hypertension, Elbow Method, Calinski-Harabasz

ABSTRAK

Meningkatnya jumlah dan ragam penyakit yang diderita masyarakat akibat perubahan gaya hidup yang dipengaruhi oleh kemajuan zaman. Salah satu Penyakit Tidak Menular (PTM) telah menjadi faktor utama dalam menyebabkan kematian di seluruh dunia. PTM yang saat ini menjadi tantangan serius dalam bidang kesehatan adalah penyakit hipertensi. Hipertensi, yang juga dikenal sebagai tekanan darah tinggi, yang merujuk pada kondisi di mana tekanan dalam arteri mengalami peningkatan. Penelitian ini berfokus pada pengelompokan dan mendiagnosa penyakit hipertensi serta bertujuan untuk melakukan klusterisasi terhadap data tersebut. Metodologi yang digunakan dalam penelitian ini adalah menggunakan CRISP-DM. Hasil dari uji coba tersebut kemudian dievaluasi menggunakan *Calinski-Harabasz* (CH). Nilai evaluasi yang dihasilkan *Calinski-Harabasz* (CH) sebesar 371.70 pada cluster (k)=3. Hasil tersebut mengindikasikan bahwa algoritma *K-Means* menghasilkan kualitas pengelompokan kluster yang optimal, yang ditunjukkan oleh nilai CH yang tinggi. Semakin tinggi nilai CH, semakin baiklah pengelompokan klusternya.

Kata Kunci : Clustering, K-Means, Hipertensi, Metode Elbow, Calinski-Harabasz

* **Korisaputri**
Email: chorisaputri@gmail.com



PENDAHULUAN

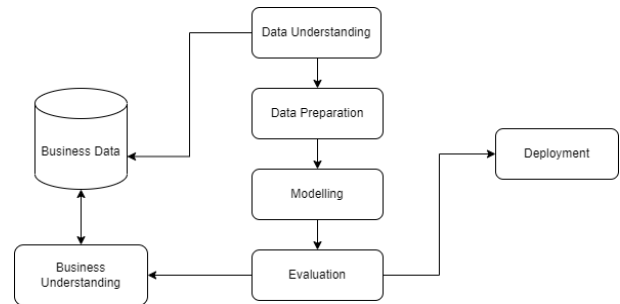
Penyakit Tidak Menular (PTM) merujuk pada kondisi kesehatan yang tidak dapat menyebar melalui kontak antar individu dan cenderung berkembang secara perlahan selama periode yang cukup lama atau (dikenal sebagai kronis)(Lisiswanti & Dananda, 2016).

Penyakit Tidak Menular (PTM) telah menjadi faktor utama dalam menyebabkan kematian di seluruh dunia. Salah satu PTM yang saat ini menjadi tantangan adalah hipertensi. Menurut *World Health Organization* (WHO), diperkirakan bahwa sekitar 22% dari total populasi dunia menderita hipertensi. Secara global, peningkatan tekanan darah menjadi salah satu faktor utama yang meningkatkan risiko kematian, menyebabkan sekitar 13% dari total kematian. Hipertensi, juga dikenal sebagai tekanan darah tinggi, merujuk pada situasi di mana tekanan darah melampaui batas normal yang biasanya adalah 120/80 mmHg(Hasanah, 2019). Tercatat di wilayah Asia Tenggara menempati peringkat ketiga dengan tingkat prevalensi hipertensi sekitar 25% dari keseluruhan populasi(Maulidah et al., 2022). Menurut informasi yang disediakan oleh Organisasi Kesehatan Dunia (WHO), terdapat sekitar 972 juta individu, atau sekitar 26,4% dari populasi global, yang menderita kondisi hipertensi. Diperkirakan bahwa angka ini berpotensi meningkat menjadi sekitar 29,2% pada tahun 2025. Angka kejadian penyakit hipertensi yang tertinggi di Jawa Tengah dapat diamati pada tiga kelompok usia khusus. Kelompok usia 31-44 tahun mencatat prevalensi sebesar 31,6%, kelompok usia 45-54 tahun memiliki prevalensi 45,3%, dan kelompok usia 55-66 tahun menunjukkan angka prevalensi tertinggi, yaitu 55,2%(Lisiswanti & Dananda, 2016). Berdasarkan temuan masalah tersebut penulis tertarik untuk melakukan diagnosis penyakit hipertensi, serta menganalisis data hipertensi yang melibatkan beberapa variabel(Siswanto et al., 2020). Tujuan penelitian ini adalah sebagai upaya untuk meningkatkan tingkat ketepatan dalam mendeteksi hipertensi dan pengelompokan hipertensi dengan *K-Means*.

METODE PENELITIAN

Studi ini memanfaatkan salah satu metode dalam data mining, yakni pengelompokan dengan algoritma *K-Means*.

Pendekatan CRISP-DM mengarahkan penelitian melalui enam langkah, dimulai dari pemahaman atas tujuan bisnis, pemahaman atas data yang ada, persiapan data, perancangan model, evaluasi, hingga penerapan hasil(Dhewayani et al., 2022).



Gambar 1 Metode CRISP-DM

Penelitian ini mengadopsi pendekatan CRISP-DM dengan langkah-langkah yang tergambar dalam diagram CRISP-DM (*Cross Industry Standard Process for Data Mining*) (Rahmadayanti et al., 2023). Dalam konteks ini, terdapat gambaran mengenai langkah-langkah tersebut yaitu:

1. *Business Understanding*: untuk memahami isu-isu yang ada dalam penelitian, yakni untuk mendukung pengelompokan serta diagnosis data pasien yang menderita hipertensi melalui penerapan metode algoritma *K-Means*.
2. *Data Understanding*: Proses pemahaman data dalam penelitian ini melibatkan pengumpulan informasi dari pasien yang menderita hipertensi. Data ini diperoleh melalui sumber-sumber seperti situs PPG-BP dan *platform* Kaggle yang memiliki kaitannya dengan data hipertensi.
3. *Data Preparation* (Pengolahan Data): Proses pengolahan data melibatkan perhitungan yang teliti. Tahap ini melibatkan langkah pembersihan data untuk mengeliminasi informasi yang tidak relevan dalam proses pengolahan.
4. *Modelling* (Pemodelan): Langkah awal dalam proses pemodelan melibatkan pemilihan *dataset* yang sesuai dengan tipe data yang akan dimanfaatkan dalam tahap eksplorasi data dengan algoritma *K-Means* sebagai metode utama dalam mengelompokkan data. Untuk pemilihan jumlah *cluster* terbaik yaitu dengan Metode *Elbow* (siku).

- 5. Evaluation (Evaluasi):** Proses untuk melakukan pengukuran hasil evaluasi dari pemodelan yang sudah diimplementasikan pada tahap modelling. Hasil evaluasi ini dilakukan untuk mengukur model yang paling baik untuk digunakan. Evaluasi yang digunakan yaitu *Calinski-Harabasz*.
- 6. Deployment (Penyebaran):** pada tahap deployment ini proses untuk membuat sebuah laporan hasil, artikel atau jurnal menggunakan hasil penelitian.

HASIL DAN PEMBAHASAN

1. Business Understanding

Tahap awal melibatkan pemahaman mendalam terhadap aspek bisnis dan pemaparan mengenai permasalahan dalam bidang penambangan data. Hal ini diperlukan agar tujuan yang diinginkan dapat tercapai secara efektif. Fokus pemahaman pada tahap ini tertuju pada dataset yang digunakan. Di tengah tingginya angka kasus hipertensi di Indonesia yang dapat menyerang berbagai rentang usia, mulai dari muda hingga lanjut usia, terdapat kebutuhan untuk mengkategorikan pasien-pasien yang menderita hipertensi.

2. Data Understanding

Dataset dalam penelitian ini merujuk pada kumpulan data tentang kondisi hipertensi pada periode 2020-2022. Dataset mengenai hipertensi ini terdiri dari 240 catatan data, yang mencakup informasi seperti ID subjek, usia (dalam tahun), jenis kelamin, tinggi dan berat badan, tekanan darah sistolik, tekanan darah diastolik, detak jantung(Heart Rate), indeks massa tubuh(BMI), diabetes, cerebral infraction, cerebrovascular disease, pendidikan, dan riwayat merokok.

| Num | Id | Sex | Age | height | weight | systolic | diastolic | Heart rate | BMI | Diabetes | Cerebral Infraction | Cerebrovascular disease | Education | Current Smoker |
|-----|-----|--------|-----|--------|--------|----------|-----------|------------|-------|----------|---------------------|-------------------------|-----------|----------------|
| 1 | 2 | Female | 45 | 152 | 63 | 160 | 89 | 97 | 27,27 | | | | 4 | 1 |
| 2 | 3 | Female | 50 | 157 | 50 | 161 | 93 | 76 | 20,28 | | | | 2 | 2 |
| 3 | 6 | Female | 47 | 150 | 47 | 101 | 71 | 79 | 20,89 | | | | 1 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | | | | ... | ... |
| 240 | 440 | Male | 36 | 160 | 57 | 110 | 67 | 67 | 20,03 | | | | 3 | 1 |

Gambar 2 Data Awal

3. Data Preparation

Pada tahap persiapan data, dilakukan langkah-langkah untuk menyesuaikan dataset agar sesuai dengan kebutuhan yang akan digunakan saat tahap pemodelan. Setelah

memperoleh dataset, beberapa langkah dilakukan dalam persiapan data, seperti berikut.

a) Data Selection

Sebanyak 240 data pasien hipertensi yang dikumpulkan antara tahun 2020-2022 telah diolah. Dari dataset ini, terdapat 14 atribut tetapi pada proses data selection ini hanya memilih 12 atribut yaitu ID pasien, jenis kelamin, usia, tinggi badan, berat badan, tekanan darah sistolik, tekanan darah diastolik, detak jantung(Heart Rate), indeks massa tubuh (BMI), pendidikan, dan riwayat merokok.

| Num | Id | Sex | Age | height | weight | systolic | diastolic | Heart rate | BMI | Diabetes | Cerebral Infraction | Cerebrovascular disease | Education | Current Smoker |
|-----|-----|--------|-----|--------|--------|----------|-----------|------------|-------|----------|---------------------|-------------------------|-----------|----------------|
| 1 | 2 | Female | 45 | 152 | 63 | 160 | 89 | 97 | 27,27 | NaN | NaN | NaN | 4 | 1 |
| 2 | 3 | Female | 50 | 157 | 50 | 161 | 93 | 76 | 20,28 | NaN | NaN | NaN | 2 | 2 |
| 3 | 6 | Female | 47 | 150 | 47 | 101 | 71 | 79 | 20,89 | NaN | NaN | NaN | 1 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | NaN | NaN | NaN | ... | ... |
| 240 | 440 | Male | 36 | 160 | 57 | 110 | 67 | 67 | 20,03 | NaN | NaN | NaN | 3 | 1 |

Gambar 3 Data Selection Awal

b) Data Preprocessing

Pada tahap pra-pemrosesan data, langkah pertama melibatkan persiapan data mentah agar siap digunakan dalam pembuatan model. Proses pra-pemrosesan ini umumnya mencakup langkah-langkah untuk membersihkan data, salah satunya adalah penanganan nilai yang hilang (*missing value*) pada beberapa fitur diantaranya yaitu Diabetes, *Cerebral Infarction*, dan *Cerebrovascular disease*.

```
data = '/kaggle/input/datarisa/dataset_awalrisa (1).csv'
df = pd.read_csv(data)
df_cleaned = df.dropna()
df_cleaned = df.dropna(axis=1)
print(df_cleaned)
```

Gambar 4 Source Code Missing Value

Dataset yang digunakan dalam penelitian ini lengkap dan bebas dari nilai yang hilang atau data yang tidak konsisten. Oleh karena itu, nantinya peneliti dapat melanjutkan ke langkah berikutnya, yaitu tahap transformasi data.

c) Data Transformation

Langkah berikutnya adalah fase transformasi data, di mana data akan diubah menjadi rentang nilai antara 0 hingga 1. Dalam

tahap ini, digunakan metode transformasi yang disebut *Standard Scaler* dengan memanfaatkan tiga fitur, yaitu Tekanan Darah Sistolik, Tekanan Darah Diastolik, dan Kebiasaan Merokok.

```
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from sklearn.preprocessing import StandardScaler

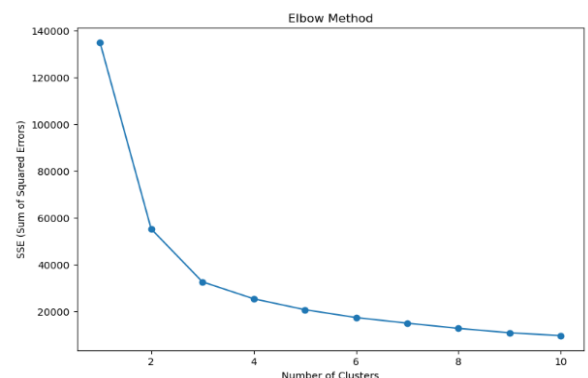
# Membaca data dari file CSV
data = '/kaggle/input/datarisa/datasetawalrisa (1).csv'
df = pd.read_csv(data)

# Menghilangkan kolom target jika ada (jika tidak digunakan dalam clustering)
X = df[['Systolic Blood Pressure(mmHg)', 'Diastolic Blood Pressure(mmHg)']] # Ganti 'target_column_name' dengan nama kolom target
sc = StandardScaler()
sc.fit(X)
X = sc.transform(X)
X
```

Gambar 5 Source Code Proses Transformasi Data

4. Modelling

Pada langkah ini, pengolahan data dilakukan dengan menerapkan teknik pengelompokan (*clustering*). Pendekatan pengelompokan yang digunakan adalah algoritma *K-Means*. Sebelum memulai proses pemodelan, langkah pertama melibatkan penentuan jumlah *cluster* yang optimal dengan memilih atribut yaitu Tekanan Darah Sistolik dan Tekanan Darah Diastolik. Untuk menentukan jumlah cluster terbaik, peneliti menggunakan metode *elbow* seperti pada Gambar 6 berikut.



Gambar 6 Grafik Method Elbow

Hasil dari grafik pada Gambar 6 mengindikasikan bahwa setelah menerapkan metode *elbow* untuk menentukan jumlah *cluster*, ditemukan bahwa jumlah *cluster* (*k*) yang optimal adalah (*k*)=3.

Setelah *cluster* ditentukan, langkah berikutnya adalah melakukan pemodelan menggunakan grafik *seaborn* pada algoritma *K-Means*. Tahapan proses pengelompokan *K-Means* pada kumpulan data pasien hipertensi diilustrasikan dalam Gambar 7 berikut.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans

# Read data from CSV file
data = '/kaggle/input/datarisa/datasetawalrisa (1).csv'
df = pd.read_csv(data)

# Assuming X contains the features you want to use for clustering
X = df[['Systolic Blood Pressure(mmHg)', 'Diastolic Blood Pressure(mmHg)', 'currentSmoker']]

kmeans = KMeans(n_clusters=4, init='k-means++', max_iter=100, n_init=10, random_state=0)
kmeans.fit(X)
centroids = kmeans.cluster_centers_

# Create a copy of X and add the 'Labels' column to the copy
X_labeled = X.copy()
X_labeled['Labels'] = kmeans.labels_

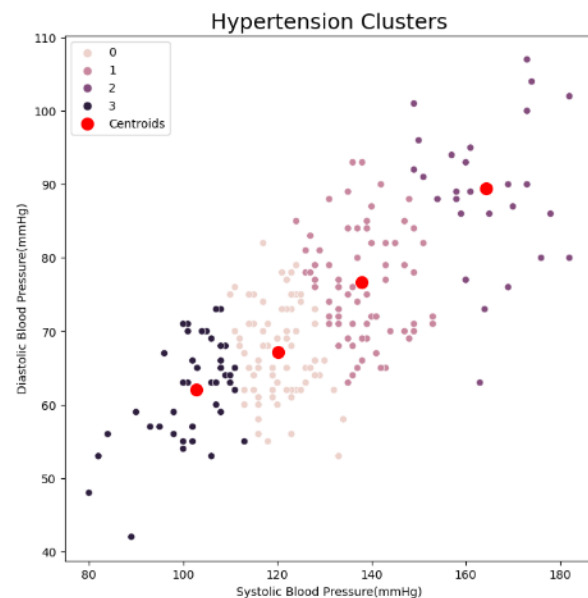
plt.figure(figsize=(8, 8))
# Use 'hue' to differentiate clusters by color
sns.scatterplot(x='Systolic Blood Pressure(mmHg)', y='Diastolic Blood Pressure(mmHg)', data=X_labeled, hue='Labels')

# Plot the centroids
plt.scatter(centroids[:, 0], centroids[:, 1], c='red', s=100, label='Centroids')
plt.title('Hypertension Clusters', fontsize=18)
plt.legend()
plt.show()
```

Gambar 7 Source Code Pemodelan K-Means Cluster

Kemudian, setelah melalui tahap pembentukan kelompok menggunakan algoritma *K-Means* sebagaimana yang terlihat pada Gambar 7, hasilnya adalah pembagian data ke dalam beberapa kelompok, sebagaimana ditunjukkan dalam Gambar 8.

Setiap anggota dalam kelompok *cluster* ini ditandai oleh titik-titik kecil dengan berbagai warna di sekitarnya, mencerminkan sebaran wilayah yang tercakup dalam setiap kelompok *cluster*.



Gambar 8 Grafik Pengelompokan Cluster

Setelah proses pemodelan *clustering* dengan menggunakan metode *k-means*, informasi mengenai hasil pengelompokan dari masing-masing data kini telah diketahui. Jumlah data yang masuk ke dalam masing-masing kelompok tampil dalam Tabel 2 sebagai berikut.

Tabel 1 Pengelompokan Tiap Cluster

| Jumlah Cluster | Jumlah Anggota Cluster |
|----------------|---|
| Cluster 0 | Normal (10 data) Pre-Hypertension (42 data) |
| Cluster 1 | Normal (29 data) Pre-Hypertension (16 data) Hypertension Stage 1 (10 data) |
| Cluster 2 | Normal (63 data) Pre-Hypertension (8 data) |
| Cluster 3 | Pre-Hypertension (2 data) Hypertension Stage 1 (49 data) Hypertension Stage 2 (11 data) |

Pengujian jumlah *Cluster* berguna untuk mengetahui jumlah *Cluster* terbaik dengan cara melihat hasil dari nilai *Calinski-Harabasz* yang dihasilkan dari pengujian yang telah dilakukan pada 240 data pasien hipertensi. Hasil pengujian dapat dilihat pada Tabel 3 berikut:

Tabel 2 Hasil Pengujian Jumlah Cluster

| Data Uji | Nilai Cluster (k) | Nilai Calinski-Harabasz |
|----------|-------------------|-------------------------|
| 240 | 2 | 343.55 |
| | 3 | 371.70 |
| | 4 | 339.50 |

Hasil pengujian jumlah *Cluster* menunjukkan masing-masing nilai yang diperoleh setiap *Cluster*, hasil nilai tertinggi didapatkan pada penggunaan nilai *Cluster* (k)=3 dengan nilai *Calinski-Harabasz* sebesar 371.70.

5. Evaluation

Setelah menyelesaikan langkah-langkah pemodelan, langkah berikutnya adalah melakukan evaluasi. Metode evaluasi yang digunakan adalah *indeks Calinski-Harabasz* dengan langkah-langkah yang tertera pada Gambar 9 di bawah ini.

```
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import calinski_harabasz_score

data = '/kaggle/input/datarisa/datasetawalrissa (1).csv'
df = pd.read_csv(data)

# Assuming X contains the features you want to use for clustering
X = df[['Systolic Blood Pressure(mmHg)', 'Diastolic Blood Pressure(mmHg)']]

# Define the range of number of clusters to test
min_clusters = 2
max_clusters = 4

# Dictionary to store CHI scores for different cluster configurations
chi_scores = {}

for n_cluster in range(min_clusters, max_clusters + 1):
    clusterer = KMeans(n_clusters=n_cluster, init='k-means++', max_iter=100, n_init=10, random_state=42)
    cluster_labels = clusterer.fit_predict(X)

    chi_score = calinski_harabasz_score(X, cluster_labels)
    chi_scores[n_cluster] = chi_score

# Print CHI scores for different cluster configurations
for n_cluster, score in chi_scores.items():
    print(f'Jumlah cluster: {n_cluster}, Nilai CHI: {score:.2f}')

# Calculate the average CHI score
average_chi_score = sum(chi_scores.values()) / len(chi_scores)
print(f'Rata-rata Nilai CHI: {average_chi_score:.2f}')
```

Gambar 9 Proses Evaluasi *Calinski-Harabasz*

```
Jumlah cluster: 2, Nilai CHI: 343.55
Jumlah cluster: 3, Nilai CHI: 371.70
Jumlah cluster: 4, Nilai CHI: 339.50
Rata-rata Nilai CHI: 351.58
```

Gambar 10 Hasil Evaluasi *Calinski-Harabasz*

Dengan demikian, jumlah *cluster* optimal terbukti berada pada *cluster* ke-3, di mana didapatkan hasil evaluasi rata-rata terbaik sebesar 371,70 menggunakan metode *Calinski-Harabasz Score*. Hal ini tercermin dari hasil pada Gambar 10 diatas, di mana skor uji evaluasi tertinggi *cluster Calinski* mencapai angka 371.70 di *cluster* (k)=3 tersebut, hal ini mengindikasikan bahwa tingkat kualitas *cluster* dalam data tersebut baik dan penggunaan indeks evaluasi *Calinski-Harabasz* menunjukkan performa yang memuaskan (Suraya & Wijayanto, 2022).

Setelah melakukan perhitungan skor evaluasi menggunakan metode *Calinski Harabasz*, langkah selanjutnya adalah mengukur tingkat akurasi model pada data yang telah dibangun dengan Adjusted Rand Index seperti yang ditunjukkan dalam Gambar 9. Dengan merujuk pada penjelasan di atas, tugas berikutnya adalah menghitung sejauh mana model yang telah dibuat mampu memberikan hasil yang akurat, dan hal ini akan dievaluasi dengan menggunakan metode yang sesuai.


```

import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import adjusted_rand_score

# Membaca data dari file CSV
data = pd.read_csv('/kaggle/input/datarisa/dataset_awsalrissa (1).csv')

# Mengambil kolom 'Systolic Blood Pressure(mmHg)' dan 'Diastolic Blood Pressure(mmHg)'
X = data[['Systolic Blood Pressure(mmHg)', 'Diastolic Blood Pressure(mmHg)']]

# Membuat objek KMeans dengan jumlah kluster yang diharapkan
kmeans = KMeans(n_clusters=3, init='k-means++', max_iter=100, n_init=10, random_state=42)

# Melakukan pengelompokan
predicted_labels = kmeans.fit_predict(X)

# Mengambil kolom 'Hypertension'
true_labels = data['Hypertension']

# Menghitung nilai Adjusted Rand Index
ari_score = adjusted_rand_score(true_labels, predicted_labels)

print("Nilai Adjusted Rand Index:", ari_score)

```

Nilai Adjusted Rand Index: 0.6563444881590754

Gambar 11 Nilai *Adjusted Rand Indeks*

Berdasarkan Gambar 11 diatas merupakan *source code* hasil perhitungan *metrics* untuk akurasi *cluster* pada dataset yang telah diolah. Oleh karena itu, hasil yang diperoleh menunjukkan tingkat ketepatan sebesar 0,656344. Dengan demikian, dapat disimpulkan bahwa akurasi dalam mendiagnosis hipertensi dapat dinyatakan cukup baik dan dapat diandalkan (Budi Dwi Satoto, Bain khusnul Khotimah, 2015).

6. Deployment

Deployment (penyebaran) mencerminkan eksekusi dari informasi yang diperoleh melalui proses pengolahan data dalam data mining. Setelah tahap evaluasi selesai, di mana hasil model dievaluasi secara menyeluruh, langkah selanjutnya adalah mengimplementasikan model yang telah dibangun. Dalam penelitian ini, implementasi dilakukan untuk menerapkan model secara praktis, dan hasilnya akan dijelaskan dalam bentuk laporan yang mengikuti langkah-langkah dalam metodologi CRISP-DM.

SIMPULAN

Penerapan metode *K-Means* untuk mengelompokkan data penyakit hipertensi menyimpulkan bahwa dalam menentukan titik optimal untuk pengelompokan, digunakan 2 fitur utama dalam data, yaitu Tekanan Darah Sistolik dan Tekanan Darah Diastolik. Metode *Elbow* digunakan untuk menentukan jumlah *cluster* yang optimal dalam algoritma *K-Means*. Penelitian diawali dengan mencari jumlah *cluster* terbaik menggunakan metode *elbow*, dimana (k)=3 adalah jumlah *cluster* terbaik. Hasil dari uji coba tersebut kemudian dievaluasi menggunakan *Calinski-Harabsz* (CH). Nilai evaluasi yang dihasilkan dengan *Calinski-Harabsz* (CH) sebesar 371.70 pada cluster (k)=3. Hasil tersebut mengindikasikan Korisaputri, Khoirun Nisa, Purwono

bahwa algoritma *K-Means* menghasilkan kualitas pengelompokan kluster yang optimal, yang ditunjukkan oleh nilai CH yang tinggi. Semakin tinggi nilai CH, semakin baiklah pengelompokan klusternya.

SARAN

Adapun saran untuk penelitian selanjutnya adalah untuk mengaplikasikan algoritma pengelompokan data (*clustering*) dengan menggunakan metode evaluasi yang beragam. Hal ini bertujuan agar hasil dari berbagai algoritma dan metode evaluasi dapat dibandingkan secara komprehensif, sehingga mampu mengidentifikasi pendekatan yang paling optimal. Dengan demikian, penelitian dapat menghasilkan analisis mendalam untuk menentukan pendekatan yang unggul.

DAFTAR PUSTAKA

- Budi Dwi Satoto, Bain khusnul Khotimah, I. (2015). Pendidikan Menggunakan Partition Around Medoids Dan. *Journal of Information System Engineering and Business Intelligence*, 1(1), 17–24.
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, 12(1), 64–77. <https://doi.org/10.34010/jati.v12i1.6674>
- Hasanah, U. (2019). Tekanan Darah Tinggi (Hipertensi). *Jurnal Keperawatan Jiwa*, 7(1), 87. <https://p2ptm.kemkes.go.id/uploads/2016/10/Tekanan-Darah-Tinggi-Hipertensi.pdf>
- Liswanti, R., & Dananda, D. N. A. (2016). Hypertension Prevention Efforts. *Majority*, 5(3), 50–54. <https://juke.kedokteran.unila.ac.id/index.php/majority/article/view/1036>
- Maulidah, K., Neni, N., & Maywati, S. (2022). Hubungan Pengetahuan, Sikap Dan Dukungan Keluarga Dengan Upaya Pengendalian Hipertensi Pada Lansia Di Wilayah Kerja Puskesmas Cikampek Kabupaten Karawang. *Jurnal Kesehatan Komunitas Indonesia*, 18(2), 484–494.

- Rahmadayanti, F., Anggraini, I., & Susanti, T. (2023). Pengklasterisasian Data Penyakit Hipertensi dengan Menggunakan Metode K-Means. *Journal of Information System Research (JOSH)*, 4(2), 737–741. <https://doi.org/10.47065/josh.v4i2.2905>
- Siswanto, Y., Widyawati, S. A., Wijaya, A. A., Salfana, B. D., & Karlina, K. (2020). Hipertensi pada Remaja di Kabupaten Semarang. *Jurnal Penelitian Dan Pengembangan Kesehatan Masyarakat Indonesia*, 1(1), 11–17. <https://doi.org/10.15294/jppkmi.v1i1.41433>
- Suraya, G. R., & Wijayanto, A. W. (2022). Perbandingan metode hierarchical clustering, k-means, k-medoids, dan fuzzy c-means dalam pengelompokan provinsi di indonesia menurut indeks khusus penanganan stunting. *Indonesian Journal of Statistics and Its Applications*, 6(2), 180–201. <https://www.stat.ipb.ac.id/journals/index.php/ijsa/article/download/899/377>