

Implementasi Algoritma K-Means Clustering dalam Penentuan Gangguan Tidur Seseorang berdasarkan Gaya Hidup

Elsa Maulida¹, Aisyah Fitri Hidayati², Febri Y Silalahi³, Yogo Dwi Prasetyo^{4*}

^{1,2,3,4} Program Studi Sistem Informasi, Institut Teknologi Telkom Purwokerto

¹21103002@ittelkom-pwt.ac.id

²21103001@ittelkom-pwt.ac.id

³21103016@ittelkom-pwt.ac.id

^{4*}yogo@ittelkom-pwt.ac.id

Abstrak— Gangguan tidur merupakan masalah kesehatan yang signifikan dan terus meningkat sehingga berdampak negatif pada kualitas hidup dan kesejahteraan individu. Gangguan tidur berkaitan erat dengan gaya hidup seseorang seperti durasi tidur sehari-hari, langkah harian, dan jumlah detak jantung yang mempengaruhi kualitas tidur. Gangguan tidur dapat mengurangi kenyamanan dan waktu tidur sehingga akan menyebabkan kualitas tidur yang buruk. Gangguan tidur yang seringkali terjadi adalah *insomnia*, *sleep apnea*, *restless legs syndrome*, *hipersomnia*, *circadian rhythm disorders*, dan *parasomnia*. Data pada penelitian ini bertujuan untuk dianalisis dan dikelompokkan berdasarkan gangguan tidur yang disebabkan oleh gaya hidup melalui salah satu metode *data mining* yaitu *K-Means Clustering* melalui identifikasi beberapa *cluster* yang memiliki gangguan tidur dan tidak memiliki gangguan tidur. Dengan menggunakan metode *K-Means* diperoleh bahwa berdasarkan data *Sleep Duration dan Quality of Sleep* terdapat 174 data pada *cluster 1* yang menunjukkan adanya gangguan tidur *insomnia*, terdapat 164 data pada *cluster 2* yang menunjukkan tidak adanya gangguan tidur, dan terdapat 36 data pada *cluster 3* yang menunjukkan adanya gangguan tidur berupa *sleep apnea* dengan struktur *silhouette score* standar (*medium structure*). Sedangkan berdasarkan data *Heart Rate dan Daily Steps* terdapat 174 data pada *cluster 1* yang menunjukkan adanya gangguan tidur *insomnia*, terdapat 164 data pada *cluster 2* yang menunjukkan adanya gangguan tidur berupa *sleep apnea*, dan terdapat 36 data pada *cluster 3* yang menunjukkan tidak adanya gangguan tidur dengan struktur *silhouette score* kuat (*strong structure*).

Kata Kunci— Gangguan tidur, gaya hidup, *data mining*, *K-Means*, *clustering*.

I. PENDAHULUAN

Kebutuhan untuk tidur merupakan sesuatu yang harus dipenuhi oleh setiap manusia untuk menghindari efek buruk seperti iritabel, depresi, letih, atau lelah dan pengendalian emosi yang buruk [1]. Gangguan tidur merupakan isu kesehatan yang signifikan, berdampak negatif pada kualitas hidup dan kesejahteraan pribadi [2]. Gangguan tidur yaitu sekelompok kondisi yang ditandai dengan ketidaksesuaian dalam kuantitas, kualitas, atau durasi tidur [2]. Gangguan tidur yang umum terjadi pada orang dewasa diantaranya *insomnia*, [3]. Gangguan tidur dapat mengurangi kenyamanan dan waktu tidur sehingga akan menyebabkan tidur yang tidak optimal [1]. Kualitas tidur didefinisikan sebagai ketika seseorang menikmati tidurnya

tanpa menunjukkan gejala gangguan tidur [1]. Di antara efek kualitas tidur yang buruk adalah penurunan aktivitas sehari-hari, rasa lelah, lemah, tanda vital tidak stabil, kondisi neurologis yang buruk, penyembuhan luka yang lambat, dan penurunan kekebalan tubuh [2].

Diperlukan penanggulangan terhadap efek negatif dari kualitas tidur yang kerap kali menyebabkan gangguan tidur. Hal tersebut dikarenakan penurunan kualitas tidur adalah fenomena yang sering ditemui dan selalu meningkat setiap tahun secara global [1]. Penanggulangan tersebut dapat dilakukan dengan menjaga aktivitas fisik seperti langkah harian yang cukup antara 4.000 hingga 18.000 langkah/hari dengan rata-rata 10.000 langkah/hari [3]. Langkah harian dapat mengukur jumlah detak jantung seseorang dengan kecepatan detak jantung yang dihasilkan antara 60 sampai 100 denyut per menit [4]. Ketiadaan aktivitas atau minimnya aktivitas fisik dapat berdampak negatif pada kualitas tidur seseorang [1]. Produksi *delta inducing peptide sleep protein* (DIPS) yang dipicu oleh kelelahan dari aktivitas fisik berkontribusi pada peningkatan kualitas tidur seseorang [1].

Berdasarkan permasalahan di atas, adanya dataset *sleep health and lifestyle* mendukung proses penerapan *clustering* untuk mengetahui beberapa *cluster* yang memiliki gangguan tidur dan tidak memiliki gangguan tidur. *Clustering* tersebut mengacu pada data *Sleep Duration dan Quality of Sleep* serta data *Heart Rate dan Daily Steps*, dengan data tersebut didapatkan dari website *Kaggle*. Klusterisasi diambil dari salah satu metode pada *data mining*. *Data mining* yaitu teknik eksplorasi dan pengkajian data ditujukan untuk melihat pola-pola yang berguna, hubungan tersembunyi, dan informasi dari dataset besar [5]. Dengan memanfaatkan berbagai teknik dan algoritma, *data mining* memungkinkan identifikasi tren, klasifikasi data, serta pembuatan prediksi yang penting bagi pengambilan keputusan [5].

Clustering merupakan teknik *data mining* yang dapat diterapkan. Dalam *data mining*, *clustering* yaitu metode untuk mencari dan mengelompokkan data menurut hal-hal serupa dari setiap data [6]. Salah satu algoritma *clustering* yang digunakan adalah *K-Means*. Algoritma *K-Means clustering* yaitu metode pengelompokan dalam analisis data dan *data mining* menggunakan pendekatan *unsupervised* berdasarkan kemiripan data, dengan cara membagi seluruh data ke dalam kelompok yang serupa dengan mengukur jarak antar data ke pusat data (*centroid*) untuk memperkirakan kesamaan [7]. Algoritma *K-*

Means memiliki implementasi yang sederhana, performa yang cepat, mudah untuk disesuaikan, dan merupakan salah satu algoritma yang paling umum digunakan [8]. Selain itu, Algoritma *K-Means* telah diterapkan dalam berbagai skenario analisis data, termasuk untuk mengkaji data kesehatan masyarakat, menjadikannya alat yang sesuai dan berguna dalam menganalisis informasi kesehatan masyarakat [8]. Penelitian terdahulu yang relevan yaitu berkaitan dengan Penerapan Algoritma *K-Means* dalam Pengkategorian Insomnia berfokus pada klasterisasi salah satu gangguan tidur yaitu insomnia memakai algoritma *K-Means Clustering* serta memperoleh hasil berupa 204 data atau sebanyak 45,13% tergolong *cluster Insomnia* dan 248 data atau sebanyak 54,87% tergolong *cluster Tidak Insomnia* [9]. Penelitian lainnya yaitu pada *Clustering* Kabupaten/Kota di Jawa Tengah Tahun 2022 berdasarkan Jumlah Kasus Kemunculan Penyakit dengan Algoritma *K-Means* yang bertujuan untuk memberikan klasterisasi pada kabupaten/kota berdasarkan data tahun 2022 mengungkapkan adanya peningkatan kasus untuk sejumlah penyakit tertentu sehingga menghasilkan *cluster* yaitu tinggi, sedang, dan rendah untuk 29 Kabupaten dan 6 Kota [10]. Pemilihan judul Algoritma *K-Means Clustering* dalam penelitian ini yaitu “Implementasi Algoritma *K-Means Clustering* dalam Penentuan Gangguan Tidur Seseorang berdasarkan Gaya Hidup” bertujuan untuk mengidentifikasi beberapa *cluster* yang memiliki gangguan tidur dan tidak memiliki gangguan tidur berdasarkan gaya hidup seseorang serta untuk mengetahui hasil terbaik dari metode yang digunakan.

II. TINJAUAN PUSTAKA

A. Clustering

Clustering yaitu cara pada *data mining* yang bekerja dengan mencari dan mengelompokkan data berdasarkan menurut hal-hal serupa dari setiap data [6]. Dalam *clustering* partisi dan objek data yang memiliki ciri serupa akan dikategorikan bersama pada satu kelompok, sementara objek data yang memiliki ciri yang berbeda akan ditempatkan pada kelompok yang berbeda pula, *Clustering* adalah salah satu metode dalam *data mining* dan telah terbukti efektif sebagai alat untuk menyelesaikan tantangan kompleks dalam ilmu komputer dan statistik [11].

Metode *clustering* terbagi menjadi dua jenis utama, terdapat dua metode pengelompokan data: *Partitioning clustering* dan *Hierarchical clustering*. Pada *Partitioning clustering*, jumlah *cluster* yang diinginkan, seperti dua atau tiga, ditentukan terlebih dahulu jumlah lainnya, sedangkan *Hierarchical clustering* merupakan metode pengelompokan data pada objek-objek yang memiliki persamaan terdekat dikelompokkan bersama secara bertahap, membentuk struktur hirarkis berbentuk pohon untuk menggambarkan tingkat kedekatan objek dari yang serupa sampai berbeda [12].

B. Algoritma K-Means Clustering

Salah satu metode analisis *cluster non-hirarkis* adalah *K-Means Clustering*, metode ini bertujuan guna membagi objek-objek ke dalam satu atau lebih kelompok berdasarkan karakteristiknya, sehingga objek dengan ciri serupa dikelompokkan dalam satu *cluster*, sementara objek dengan ciri yang berbeda dikelompokkan dalam *cluster* lain [12]. Algoritma *K-Means* memiliki keunggulan dalam hal kemudahan implementasi, tingkat konvergensi yang cepat, dan kemampuan untuk menghasilkan *cluster* yang lebih rapat dibandingkan dengan metode hirarki, namun, algoritma ini peka terhadap pemilihan pusat *cluster* awal dan sulit menentukan jumlah *cluster* yang ideal. [J8]. Metode *K-Means* menggunakan k pola untuk titik awal *centroid* secara acak. Posisi awal *centroid* yang dipilih secara acak akan mempengaruhi hasil iterasi dalam menentukan *centroid cluster*. Ketika *centroid* baru tetap di posisi, nilai K dipilih sebagai titik awal ditentukan dengan menggunakan rumus *Euclidean Distance*, yang menentukan jarak terdekat antara *centroid* dan data/objek. Data yang paling dekat dengan *centroid* akan membentuk *cluster* [13]. Tata cara dari perhitungan Algoritma *K-Means* adalah sebagai berikut [14] [15] [11] [16].

1. Menentukan nilai k sebagai jumlah *cluster* yang akan dibentuk menggunakan rumus elbow criterion dengan rumus sebagai berikut :

$$SSE = \sum_{K=1}^K \sum_{x_i=3k} \|N_i C_k\| \quad (1)$$

Keterangan :

x = data

c = *centroid* atau pusat data = *cluster*

2. Menentukan titik k pusat *cluster (centroid)* awal yang dilakukan secara acak. *Centroid* awal ditentukan secara acak dengan data objek sebanyak k *cluster*, lalu menghitung *centroid cluster* ke- i selanjutnya, menggunakan rumus :

$$v = \frac{\sum_{i=1}^n x_i}{n} : i = 1, 2, 3, \dots, n \quad (2)$$

V = *centroid* pada *cluster*

x_i = objek ke- i

n = jumlah objek yang menjadi anggota *cluster*

3. Hitung jarak masing-masing objek ke masing-masing *centroid* dari masing-masing *cluster* menggunakan rumus *Euclidean Distance* sebagai berikut :

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} : i = 1, 2, 3, \dots, n \quad (3)$$

Keterangan :

d = jarak antara x dan y

x = data pusat *cluster*

y = data pada atribut

i = setiap data

n = jumlah data

x_i = data pada pusat *cluster* ke i

y_i = data pada setiap data ke i

4. Tempatkan masing-masing objek ke dalam *centroid* terdekat. Perpindahan objek ke dalam masing-masing *cluster* saat iterasi secara umum dilakukan melalui *hard k-means* yang mana ditegaskan setiap objek dinyatakan sebagai bagian dari *cluster* dengan mengukur kedekatan jarak sifatnya terhadap titik pusat *cluster*.

5. Melakukan iterasi kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan.

6. Lakukan langkah ketiga apabila posisi *centroid* baru tidak sama.

C. Data Mining

Data Mining adalah teknik untuk menganalisis pola dan karakteristik di masa yang akan datang serta menggali informasi tak terduga yang belum pernah ditemukan sebelumnya dari basis data besar [17]. *Data mining* atau yang juga dikenal sebagai penemuan pengetahuan dalam basis data (KDD) adalah proses mengumpulkan, mengolah, dan menerapkan data berukuran besar untuk mengidentifikasi aturan atau pola yang dapat disimpan dalam basis data atau media penyimpanan lainnya [7]. *Data mining* bertujuan utama untuk mengolah data di dalam basis data sehingga mendapatkan wawasan baru yang bermanfaat [13].

D. Silhouette Score

Silhouette Score digunakan untuk mengevaluasi seberapa baik suatu objek ditempatkan dalam cluster, mencerminkan kualitas serta kekuatan *cluster* tersebut [18]. *Silhouette score* memiliki rentang nilai dari -1 hingga 1, di mana semakin tinggi nilai tersebut menunjukkan semakin baik kualitas *clustering* [19]. Rentang nilai *silhouette Score* dituliskan sebagai berikut [20].

1. $0.7 < SC \leq 1$ Struktur yang kuat (*strong structure*).
2. $0.5 < SC \leq 0.7$ Struktur yang standar (*medium structure*).
3. $0.25 < SC \leq 0.5$ Struktur yang lemah (*weak structure*).
4. $SC \leq 0.25$ Tidak memiliki struktur (*no structure*).

E. Gangguan Tidur

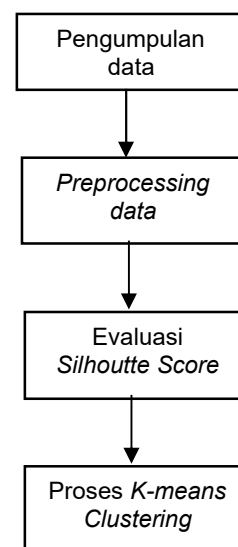
Gangguan tidur mencakup kondisi di mana terdapat fluktuasi dalam waktu, kualitas, atau jumlah tidur seseorang yang dapat mengganggu fungsi normal atau pola hidup dengan beragam penyebab mencakup masalah medis jangka panjang atau jangka pendek, serta kebiasaan tidur yang tidak sehat, stres, dan lingkungan yang mempengaruhi kualitas hidup dan jika diperkirakan gangguan tidur timbul karena masalah mental maupun fisik, perlu dianggap sebagai gangguan kesehatan yang membutuhkan penanganan [21].

F. Gaya Hidup

Gaya hidup adalah pola kehidupan seseorang, termasuk aktivitas, minat, dan sikap yang mereka miliki yang diekspresikan dalam berbagai aspek kehidupan [22].

III. METODOLOGI PENELITIAN

Penelitian ini menerapkan algoritma *K-means* untuk *clustering*, dengan data kasus implementasi algoritma *K-means clustering* digunakan untuk menentukan gangguan tidur seseorang berdasarkan gaya hidup. Data yang dipergunakan diambil dari *Kaggle* dengan judul dataset "*Sleep Health and Lifestyle Dataset*". Penentuan jumlah *cluster* dilakukan melalui metode *elbow* dengan tujuan untuk mengoptimalkan jumlah *cluster* [23]. Dataset yang telah dikumpulkan kemudian diolah. Adapun Gambar 1 menampilkan tahapan-tahapan penelitian yang akan dilaksanakan.



Gbr. 1 Diagram alir penelitian

Pada Gambar 1 ditunjukkan bahwa alur penelitian terdiri dari pengumpulan data, *preprocessing*, proses *K-Means clustering*, dan evaluasi *Silhouette Score*. Adapun penjabaran dari gambar tahapan penelitian yaitu sebagai berikut:

1. **Pengumpulan Data**
Pada tahap ini, data yang digunakan untuk penelitian didapat dari *Kaggle*, tepatnya dataset "*Sleep Health and Lifestyle Dataset*". Dataset ini mencakup berbagai informasi mengenai kebiasaan tidur dan gaya hidup seseorang. Mengumpulkan semua data yang relevan untuk analisis lebih lanjut. Dataset ini akan menjadi basis untuk seluruh proses analisis *clustering* yang akan dilakukan.

2. Preprocessing Data

Tahap ini melibatkan pembersihan dan transformasi data untuk memastikan bahwa data siap untuk dianalisis. Proses ini mungkin meliputi pembersihan data (menangani nilai yang hilang, menghapus duplikat, dll.), menormalkan atau menstandarisasi data, dan mungkin mengubahnya ke format

yang lebih sesuai. Tujuannya adalah untuk menyiapkan dataset yang dapat dianalisis secara efisien dan efektif.

3. Proses *K-Means Clustering*

Pada tahap ini akan dilakukan proses *K-Means Clustering*. Tahap ini melibatkan penggunaan algoritma *K-Means* untuk melakukan *clustering* atau pengelompokan data. Adapun proses *K-Means Clustering* dapat dilakukan dengan langkah berikut:

a. Menentukan *n-Cluster*

Di tahap ini, peneliti menentukan jumlah *cluster* yang optimal untuk digunakan dalam analisis. Metode yang digunakan untuk menentukan jumlah *cluster* pada penelitian ini yaitu dengan menggunakan metode *Elbow* yang menggambarkan total dalam *cluster sum of square* (WCSS) terhadap jumlah *cluster*, dan mencari titik "siku" yang menandakan jumlah *cluster* yang optimal. Tujuannya adalah untuk menentukan jumlah *cluster* yang paling sesuai untuk analisis *K-Means Clustering*, sehingga hasilnya dapat diinterpretasikan dengan baik.

b. Melakukan Visualisasi *Cluster*

Setelah menentukan jumlah *cluster* dan menjalankan algoritma *K-Means*, langkah selanjutnya adalah memvisualisasikan hasil *clustering*. Teknik visualisasi diterapkan untuk menafsirkan dan memahami *cluster*. Tujuannya adalah untuk membantu dalam interpretasi hasil *clustering* dan memahami karakteristik dari setiap *cluster* yang terbentuk.

c. Menentukan Grafik Jumlah Anggota *Cluster*

Menentukan grafik jumlah anggota *cluster* ini sangat membantu dalam memahami bagaimana data terdistribusi di antara *cluster-cluster* yang berbeda. Misalnya, ukuran *cluster* dapat memberikan wawasan tentang prevalensi pola tidur tertentu dalam populasi yang diteliti. Selain itu, mengetahui jumlah anggota dalam setiap *cluster* juga dapat membantu dalam analisis lebih lanjut. Grafik jumlah anggota *cluster* memberikan gambaran yang jelas tentang distribusi data dan membantu dalam interpretasi hasil *clustering*, yang pada gilirannya dapat digunakan untuk menarik kesimpulan yang lebih mendalam dan membuat keputusan yang lebih informasional dalam penelitian mengenai gangguan tidur dan seseorang berdasarkan gaya hidup.

Dengan melakukan proses *K-Means Clustering* di atas, *K-Means* membantu untuk memudahkan visualisasi dan analisis lebih lanjut.

4. Evaluasi *Silhouette Score*

Langkah terakhir adalah mengevaluasi kualitas *clustering* dengan menggunakan *Silhouette Score*. *Silhouette Score* mengukur seberapa mirip suatu objek dengan *clusternya* sendiri dibandingkan dengan *cluster* lain, memberikan wawasan tentang kekompakan, dan pemisahan *cluster*. Skor *Silhouette* yang tinggi menunjukkan *cluster* yang terdefinisi dengan baik dan terpisah dengan baik. Tujuan mengevaluasi

Silhouette Score yaitu untuk menilai seberapa baik *cluster* yang dihasilkan oleh algoritma *K-Means* dan memastikan bahwa *cluster* yang terbentuk memiliki kualitas yang baik dan interpretasi yang jelas.

Dengan mengikuti tahapan-tahapan ini, penelitian ini mampu mengimplementasikan algoritma *K-Means Clustering* untuk menentukan gangguan tidur berdasarkan gaya hidup dengan cara yang terstruktur dan ilmiah.

IV. HASIL DAN PEMBAHASAN

A. Dataset Awal

Data yang digunakan dalam dataset adalah *data Sleep Duration, Quality of Sleep, Heart Rate, dan Daily Steps*. Data diperoleh dari *website Kaggle* dengan nama dataset *Sleep Health and Lifestyle Dataset*. Dataset awal penelitian ditunjukkan pada tabel 1.

TABEL I
 DATASET AWAL BAGIAN HEAD

Dataset Awal			
<i>Sleep Duration</i>	<i>Quality Of Sleep</i>	<i>Heart Rate</i>	<i>Daily Steps</i>
6.1	6	77	4200
6.2	6	75	10000
6.2	6	75	10000
5.9	4	85	3000
5.9	4	85	3000
5.9	4	82	3500
7.8	7	70	8000
7.8	7	70	8000
7.8	7	70	8000

Berdasarkan data di atas, *clustering* dilakukan berdasarkan kolom keempat dan kelima yang menunjukkan data *Sleep Duration* dan *Quality of Sleep*. Proses *clustering* diawali dengan memanggil dataset dan memberikan nama pada variabel X untuk menampilkan data keempat dan kelima.

B. Preprocessing Data

Kemudian, dilakukan *scaling* data menggunakan proses *Preprocessing MinMaxScaler*. Berikut merupakan data sebelum dilakukan *scaling*.

```
print("Data sebelum scaling:")
print(df)

↔ Data sebelum scaling:
   0  6.1  6.0
   1  6.2  6.0
   2  6.2  6.0
   3  5.9  4.0
   4  5.9  4.0
  ..  ...  ...
369  8.1  9.0
370  8.0  9.0
371  8.1  9.0
372  8.1  9.0
373  8.1  9.0

[374 rows x 2 columns]
```

Gbr. 2 Data *sleep duration* dan *quality of sleep* sebelum *scaling*

```
print("\nData setelah scaling:")
print(scaled_df)

↔ Data setelah scaling:
   0  0.111111  0.4
   1  0.148148  0.4
   2  0.148148  0.4
   3  0.037037  0.0
   4  0.037037  0.0
  ..  ...  ...
369  0.851852  1.0
370  0.814815  1.0
371  0.851852  1.0
372  0.851852  1.0
373  0.851852  1.0

[374 rows x 2 columns]
```

Gbr. 4 Data *sleep duration* dan *quality of sleep* setelah *scaling*

```
print("Data sebelum scaling:")
print(df)

↔ Data sebelum scaling:
   0  77  4200
   1  75 10000
   2  75 10000
   3  85  3000
   4  85  3000
  ..  ..  ...
369  68  7000
370  68  7000
371  68  7000
372  68  7000
373  68  7000

[374 rows x 2 columns]
```

Gbr. 3 Data *heart rate* dan *daily steps* sebelum *scaling*.

```
print("\nData setelah scaling:")
print(scaled_df)

↔ Data setelah scaling:
   0  0.571429  0.171429
   1  0.476190  1.000000
   2  0.476190  1.000000
   3  0.952381  0.000000
   4  0.952381  0.000000
  ..  ...  ...
369  0.142857  0.571429
370  0.142857  0.571429
371  0.142857  0.571429
372  0.142857  0.571429
373  0.142857  0.571429

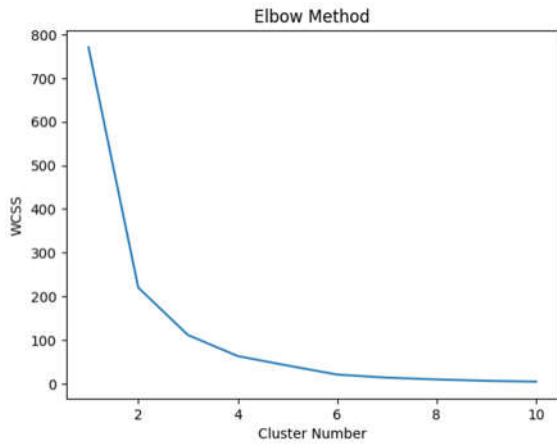
[374 rows x 2 columns]
```

Gbr. 5 Data *heart rate* dan *daily steps* setelah *scaling*

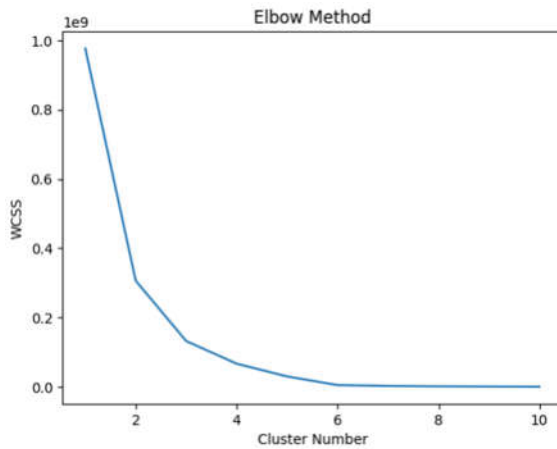
Setelah mengetahui data sebelum *scaling*, dapat dilihat bahwa terdapat jarak antar data. Sehingga selanjutnya dilakukan *scaling* data untuk mengubah skala data dalam dataset dengan skala yang serupa sehingga data dan menghasilkan rentang data baru sebagai berikut.

C. Menentukan *n-Cluster*

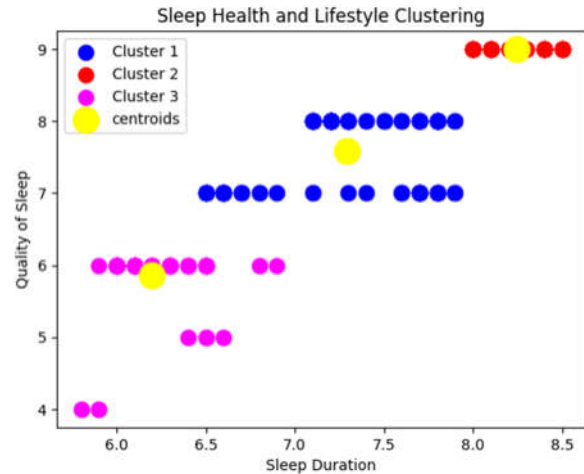
Data yang telah melalui *Preprocessing*, dicari jumlah *n-kluster* melalui *Elbow Method* dengan melihat *plot WCSS* dan jumlah *cluster*. *Plot WCSS* digunakan untuk menambahkan nilai *inertia* sehingga dihasilkan jumlah *cluster* yaitu 3. Berikut merupakan *Elbow Method* pada data *Sleep Duration* dan *Quality of Sleep* serta data *Heart Rate* dan *Daily Steps*.



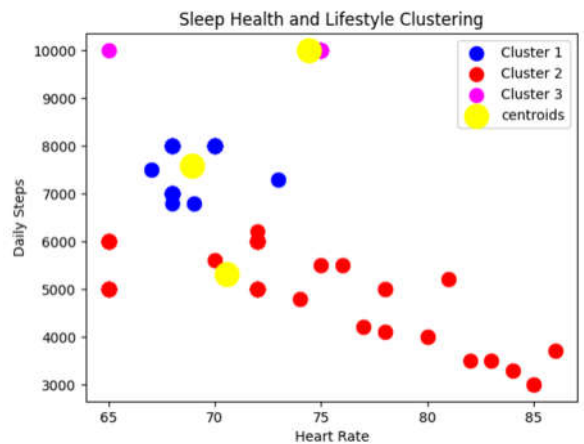
Gbr. 6 Elbow method sleep duration dan quality of sleep



Gbr. 7 Elbow method heart rate dan daily steps



Gbr. 8 Visualisasi cluster sleep duration dan quality of sleep



Gbr. 9 Visualisasi cluster heart rate dan daily steps

Grafik yang dihasilkan *elbow method* di atas menunjukkan *n-cluster* berbentuk siku pada angka 3.

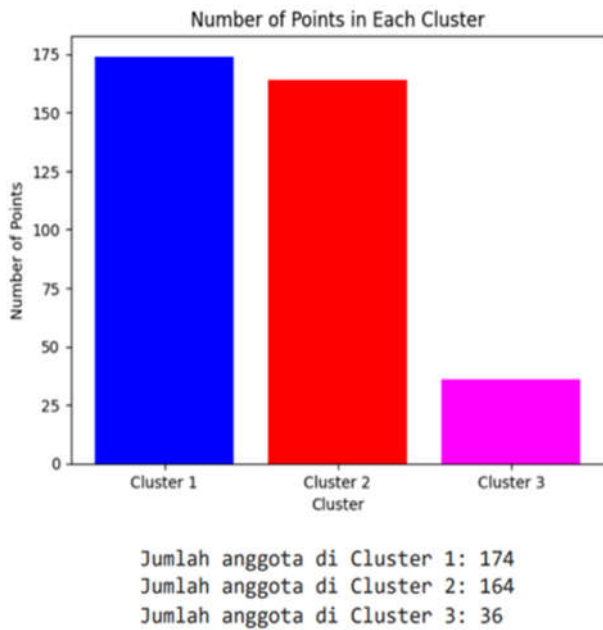
D. Melakukan Visualisasi Cluster

Selanjutnya, data dikelompokkan dalam X menjadi 3 *cluster* sehingga menghasilkan label *cluster* yang menunjukkan titik-titik *cluster*. Hasil plot *cluster* divisualisasikan menggunakan *matplotlib.pyplot* sebagai berikut.

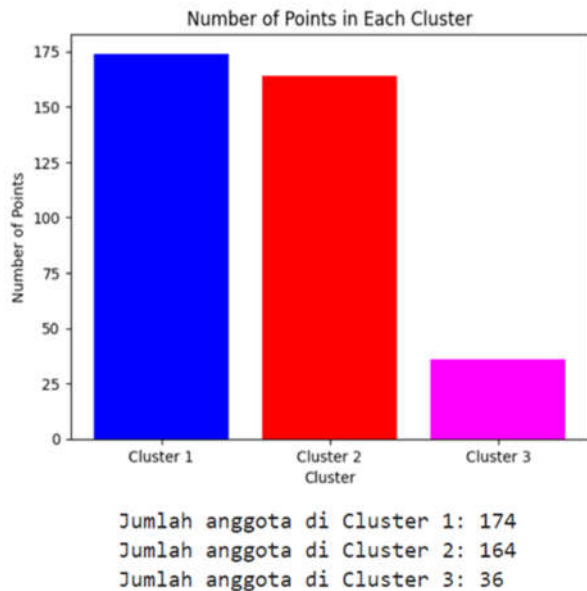
Visualisasi cluster dibuat untuk mendistribusi data ke dalam *cluster* dan mengevaluasi hasil *clustering* secara visual.

E. Evaluasi Silhouette Score

Setelah *cluster* di visualisasikan, perlu dilihat kembali distribusi data pada tiap *cluster* untuk mengetahui jumlah anggota dari tiap *cluster* tersebut. Grafik akan memudahkan dalam identifikasi *cluster* sehingga penyampaian hasil *clustering* akan lebih jelas. Berikut merupakan jumlah anggota pada tiap *cluster*.



Gbr.10 Jumlah anggota cluster sleep duration dan quality of sleep



Gbr. 11 Jumlah anggota cluster heart rate dan daily steps

F. *Evaluasi Silhouette Score*

Setelah itu, kualitas *clustering* dievaluasi menggunakan matrik evaluasi *silhouette score* untuk mengukur seberapa baik setiap titik data yang sudah dikelompokkan dengan sesama anggota sama. Pengukuran dengan *silhouette score* dilakukan dengan melihat rentang *cluster* kedua hingga ke-10 melalui iterasi maksimal sebanyak 50 iterasi. Berikut merupakan hasil perhitungan *silhouette score*.

```
range_n_clusters = [2, 3, 4, 5, 6, 7, 8, 9, 10]
for num_clusters in range_n_clusters:
    #initials KMeans
    kmeans = KMeans(n_clusters = num_clusters, max_iter=50)
    kmeans.fit(X)
    cluster_labels = kmeans.labels_

    #Silhouette Score
    silhouette_avg = silhouette_score(X, cluster_labels)
    print("For n_clusters={0}, the silhouette score is{1}".format(num_clusters, silhouette_avg))

    For n_clusters=2, the silhouette score is0.6262582195494373
    For n_clusters=3, the silhouette score is0.6319914771456847
    For n_clusters=4, the silhouette score is0.6992561829412737
    For n_clusters=5, the silhouette score is0.7467109451774713
    For n_clusters=6, the silhouette score is0.7808433980859492
    For n_clusters=7, the silhouette score is0.79654534340779
    For n_clusters=8, the silhouette score is0.807864361258121
    For n_clusters=9, the silhouette score is0.771924909207629
    For n_clusters=10, the silhouette score is0.759645826303865
```

Gbr. 12 Hasil silhouette score sleep duration dan quality of sleep

```
For n_clusters=2, the silhouette score is0.6483415394384682
For n_clusters=3, the silhouette score is0.7233444078004578
For n_clusters=4, the silhouette score is0.7468071813673947
For n_clusters=5, the silhouette score is0.8877413678062895
For n_clusters=6, the silhouette score is0.9388125223159455
For n_clusters=7, the silhouette score is0.9401665351750614
For n_clusters=8, the silhouette score is0.9528057838727823
For n_clusters=9, the silhouette score is0.9580422287689705
For n_clusters=10, the silhouette score is0.9597802800750097
```

Gbr. 13 Hasil silhouette score heart rate dan daily steps

V. KESIMPULAN

Berdasarkan proses *clustering*, diperoleh hasil klasterisasi sebagai berikut. Diperoleh tiga pengkategorian gangguan tidur pada klasterisasi menggunakan algoritma *K-Means*, yaitu *sleep apnea*, *insomnia*, dan tidak ada gangguan tidur. Diperoleh bahwa pada data *Sleep Duration* dan *Quality of Sleep* terdapat 174 data pada *cluster* 1 yang menunjukkan adanya gangguan tidur *insomnia*, terdapat 164 data pada *cluster* 2 yang menunjukkan tidak adanya gangguan tidur, dan terdapat 36 data pada *cluster* 3 yang menunjukkan adanya gangguan tidur berupa *sleep apnea*. Diperoleh bahwa pada data *Heart Rate* dan *Daily Steps* terdapat 174 data pada *cluster* 1 yang menunjukkan adanya gangguan tidur *insomnia*, terdapat 164 data pada *cluster* 2 yang menunjukkan adanya gangguan tidur berupa *sleep apnea*, dan terdapat 36 data pada *cluster* 3 yang menunjukkan tidak adanya gangguan tidur. Hasil pengujian *silhouette score* pada data *Sleep Duration* dan *Quality of Sleep* dapat disimpulkan bahwa pemilihan jumlah *cluster* sebanyak 3 menghasilkan skor sebesar 0.6319914771456847 menunjukkan struktur yang standar (*medium structure*). Hasil pengujian *silhouette score* pada data *Heart Rate*, dan *Daily Steps* dapat disimpulkan bahwa pemilihan jumlah *cluster* sebanyak 3 menghasilkan skor sebesar 0.7233444078004578 menunjukkan struktur yang kuat (*strong structure*).

VI. REFERENSI

- [1] I. G. Wiyana, Y. Dahlia, N. P. Rinaya, and D. Utary, "Hubungan Indeks Massa Tubuh, Aktivitas Fisik dan Durasi Penggunaan Media Sosial dengan Kualitas Tidur Siswa," *J. Glob. Ilm.*, vol. 1, no. 2, pp. 130–142, 2023.
- [2] W. K. Rohmah and Y. D. P. Santik, "Determinan Kualitas Tidur Pada Santri di Pondok Pasantren," *Higeia J. Public Heal. Res. Dev.*, vol. 4, no. 3, pp. 649–659, 2020.
- [3] C. Tudor-Locke, J. E. Williams, J. P. Reis, and D. Pluto, "How Many Steps/day are Enough? For Adults," *Sport. Med.*, vol. 32, no. 12, pp. 795–808, 2002.
- [4] D. R. Pradana, M. I. Sari, and D. Rimasa, "Monitoring Denyut Nadi Dan

- Suhu Tubuh Menggunakan Pulse Sensor Dan Sensor Suhu Pada Atlet Anggar Koni Kota Bandung,” *e-Proceeding Appl. Sci.*, vol. Vol. 9, no. No. 6, pp. 3042–3048, 2023.
- [5] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” *J. Ilm. Inform. Dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024.
- [6] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, “Analisis K-Means Clustering pada Data Sepeda Motor,” *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020.
- [7] A. Rohmah, “Impelementasi Algoritma K-Means Clustering Analisis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus : SMK Yaspmi Gegerbitung),” *J. Rekayasa Teknol. Nusa Putra*, vol. 4, no. 2, pp. 30–35, 2021.
- [8] S. U. Tarigan, S. Saniman, and M. Yetri, “Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat Menggunakan Algoritma K-Means,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 3, p. 193, 2022.
- [9] S. Informasi and P. B. Kaltara, “Penerapan Algoritma K-Means Dalam Pengkategorian Insomnia,” vol. 8, no. 1, pp. 117–122, 2024.
- [10] N. A. Mufid, D. R. N. Hanum, and A. H. Sidiq, “Clustering Kabupaten/Kota di Jawa Tengah Tahun 2022 berdasarkan Jumlah Kasus Kemunculan Penyakit dengan Algoritma K-Means,” *J. Pendidik. Mat.*, vol. 1, no. 1, p. 14, 2023.
- [11] Gustiendina, Mh. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019.
- [12] N. K. Zuhal, “Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering,” *Pros. Semin. Nas. Teknol. dan Sains*, vol. 1, pp. 200–205, 2022.
- [13] F. Nurdyansyah and I. Akbar, “Implementasi Algoritma K-Means untuk Menentukan Persediaan Barang pada Poultry Shop,” *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 2, pp. 86–94, 2021.
- [14] R. Swastika, S. Mukodimah, F. Susanto, M. Mushlihudin, and S. Ipnuwati, *Implementasi Data Mining (Clustering, Association, Prediction, Estimation, Classification)*, Digital. Indramayu, Jawa Barat: Adab CV. Adanu Abimata, 2023.
- [15] R. Rahmattullah, Indwiarti Indwiarti, and A. A. Rohmawati, “Clustering Harga Rumah: Perbandingan Model K-Means dan Gaussian Mixture Model,” *e-Proceeding Eng.*, vol. 10, no. 3, pp. 3441–3449, 2023.
- [16] A. K. K-medoids, “Penerapan Data Mining Clustering Dalam Menentukan Tingkat Pembelian Kredit Tertinggi,” vol. IV, pp. 134–138, 2023.
- [17] S. Setyaningtyas, B. I. Nugroho, and Z. Arif, “Tinjauan Pustaka Sistematis pada Data Mining: Studi Kasus Algoritma K-Means Clustering,” *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022.
- [18] D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019.
- [19] M. Hernita, E. Pramesty, Y. U. Aprilia, R. Bryan, J. Purba, and M. Athoillah, “Optimalisasi Kualitas Pendidikan Provinsi Nusa Tenggara Timur: Klasterisasi Kabupaten Dan Kota Menggunakan Algoritma K-Means Dengan Metode Elbow Dan Shilluette Score,” *Semin. Nas. Has. Ris. Dan Pengabd.*, pp. 359–368, 2023.
- [20] Zaini Miftach, “Menentukan Jumlah Cluster Terbaik Pada K-Mean Untuk Jumlah Data Terjangkit Covid-19,” pp. 53–54, 2018.
- [21] U. Faridah, D. Kusumawati, S. Rahayu, and D. Wahab, “Hubungan Kebiasaan Mengonsumsi Kopi Dengan Gejala Gangguan Tidur Pada Lansia di Desa Tempuran Demak 2018,” pp. 228–241, 2021.
- [22] N. S. Azizah, “Pengaruh Literasi Keuangan, Gaya Hidup Pada Perilaku Keuangan Pada Generasi Milenial,” *Prism. (Platform Ris. Mhs. Akuntansi)*, vol. 01, no. 02, pp. 92–101, 2020.
- [23] 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.