

# Menjelajahi Wawasan Industri Musik: Klasterisasi Lagu Terpopuler Di *Spotify* 2024 Dengan Metode *K-Means Clustering*

Sri Erina Damayanti<sup>1</sup>, Mohamad Ilham Fajriana<sup>2</sup>, Dea Meilani<sup>3</sup>, Shofi Siti Fatimah<sup>4</sup>

Teknik Informatika  
Universitas Teknologi Bandung  
Bandung, Indonesia

<sup>1</sup>[srikandibrilliant1@gmail.com](mailto:srikandibrilliant1@gmail.com), <sup>2</sup>[iruhamfajriana9d18@gmail.com](mailto:iruhamfajriana9d18@gmail.com), <sup>3</sup>[deameilani2022@gmail.com](mailto:deameilani2022@gmail.com),

<sup>4</sup>[shofisf210@gmail.com](mailto:shofisf210@gmail.com)

Correspondence : e-mail: iruhamfajriana9d18@gmail.com

Diajukan: 27 Agustus 2024; Direvisi: 28 Agustus 2024; Diterima: 29 Agustus 2024

## Abstrak

Penelitian ini mengeksplorasi dinamika industri musik dengan mengklasterisasi lagu-lagu terpopuler di *Spotify* tahun 2024 menggunakan metode *K-Means Clustering*. Latar belakang penelitian ini adalah perubahan signifikan dalam cara musik diakses dan dinikmati oleh konsumen, terutama melalui platform streaming seperti *Spotify*. Masalah yang akan diselesaikan adalah identifikasi tren musik dan faktor-faktor yang mempengaruhi popularitas lagu. Data dari dataset "*Most Streamed Spotify Songs 2024*" dianalisis menggunakan metode *K-Means Clustering* untuk mengelompokkan lagu berdasarkan karakteristik audio mereka. Hasil penelitian ini menunjukkan lima cluster utama dengan karakteristik unik masing-masing, yang mencerminkan pola popularitas dan preferensi musik global. Temuan ini penting untuk pemahaman lebih dalam tentang dinamika industri musik saat ini dan dapat memberikan wawasan strategis bagi para pelaku industri dalam mengembangkan strategi promosi dan distribusi yang lebih efektif.

**Kata kunci:** *Spotify*, *K-Means Clustering*, popularitas lagu, tren musik, analisis data.

## Abstract

This research explores the dynamics of the music industry by clustering the most popular songs on *Spotify* in 2024 using the *K-Means Clustering* method. The background of this study is the significant change in how music is accessed and enjoyed by consumers, especially through streaming platforms like *Spotify*. The problem to be solved is the identification of music trends and factors influencing song popularity. Data from the "*Most Streamed Spotify Songs 2024*" dataset were analyzed using the *K-Means Clustering* method to group songs based on their audio characteristics. The results of this study reveal five main clusters, each with unique characteristics that reflect global music popularity patterns and preferences. These findings are crucial for a deeper understanding of the current dynamics of the music industry and can provide strategic insights for industry players in developing more effective promotion and distribution strategies.

**Keywords:** *Spotify*, *K-Means Clustering*, song popularity, music trends, data analysis.

## 1. Pendahuluan

Menurut Sylado, seni musik adalah bentuk perwujudan yang hidup dari sebuah kumpulan ilusi dan lantunan suara penciptanya menggunakan alunan musik dengan nada yang berjiwa dan bisa menggerakkan isi hati para pendengarnya. Musik telah menjadi bagian integral dari kehidupan manusia, memainkan peran penting dalam berbagai aspek seperti hiburan, budaya, dan ekspresi emosional. Seiring dengan perkembangan teknologi digital, cara kita mengakses dan menikmati musik telah mengalami transformasi signifikan. *Spotify*, sebagai salah satu layanan musik digital, *podcast*, dan video terbesar di dunia, memberikan akses kepada pengguna untuk menikmati jutaan lagu dan konten lain dari kreator di seluruh dunia. Platform ini telah mengubah lanskap industri musik, memungkinkan pengguna untuk mengakses jutaan lagu secara instan. Perubahan ini tidak hanya berdampak pada cara musik didistribusikan, tetapi juga pada bagaimana popularitas lagu diukur dan dipahami.

Perkembangan teknologi digital juga membawa dampak signifikan pada perilaku konsumen dalam menikmati musik. Jika dahulu orang harus membeli album fisik atau mendengarkan radio untuk menikmati lagu-lagu favorit mereka, kini dengan adanya layanan streaming seperti *Spotify*, akses terhadap musik menjadi lebih mudah dan cepat. Hal ini juga mengubah cara industri musik beroperasi, dimana popularitas lagu tidak lagi diukur berdasarkan penjualan album fisik, melainkan jumlah streaming dan interaksi pengguna di *platform* digital.

*Dataset* "Lagu *Spotify* Paling Banyak Didengarkan Tahun 2024" memberikan gambaran komprehensif tentang preferensi musik global. Data ini mencakup informasi detail mengenai lagu-lagu yang paling banyak didengarkan di *Spotify* sepanjang tahun 2024, termasuk nama lagu, artis, tanggal rilis, jumlah streaming, dan popularitas di berbagai *platform* musik lainnya. Dengan menganalisis *dataset* ini, kita dapat mengidentifikasi tren musik, pola popularitas, dan karakteristik lagu yang dominan pada tahun 2024. Analisis ini penting untuk memahami dinamika industri musik saat ini dan bagaimana perubahan teknologi mempengaruhi preferensi musik konsumen.

Fokus penelitian ini adalah mengidentifikasi dan memahami tren serta pola popularitas lagu di *Spotify* pada tahun 2024 melalui analisis data. Pendekatan yang digunakan adalah analisis statistik dan pemodelan data untuk mengekstraksi informasi berharga dari *dataset* tersebut. Metode *K-means Clustering* digunakan untuk mengelompokkan lagu-lagu berdasarkan karakteristik audio mereka, membantu mengidentifikasi tren musik dan faktor-faktor yang mempengaruhi popularitas lagu.

Hasil yang diharapkan dari penelitian ini adalah pemahaman yang lebih mendalam tentang preferensi musik global dan pola popularitas lagu di *Spotify*. Penelitian ini juga diharapkan dapat memberikan kontribusi terhadap literatur musik digital dan industri musik, serta membantu dalam pengambilan keputusan strategis yang lebih baik di masa depan. Dengan demikian, studi ini dapat membuka peluang untuk penelitian lebih lanjut mengenai pengaruh *platform* streaming musik terhadap industri musik secara keseluruhan. Penelitian ini juga diharapkan dapat memberikan rekomendasi bagi para pelaku industri musik dalam mengembangkan strategi promosi dan distribusi yang lebih efektif, serta membantu artis dan produser dalam menciptakan karya-karya yang lebih sesuai dengan preferensi pasar.

Penelitian ini tidak hanya penting bagi industri musik, tetapi juga bagi akademisi dan peneliti yang tertarik pada bidang musik digital dan analisis data. Dengan memahami tren dan pola popularitas lagu di *Spotify*, kita dapat mendapatkan wawasan yang lebih luas tentang bagaimana teknologi mempengaruhi industri kreatif dan perilaku konsumen. Selain itu, penelitian ini juga dapat menjadi dasar bagi studi-studi lanjutan yang mengeksplorasi berbagai aspek lain dari musik digital, seperti dampak sosial dan budaya dari *platform* streaming, serta implikasi ekonomi dari perubahan dalam industri musik.

## 2. Metode Penelitian

Penelitian ini menggunakan metode analisis kuantitatif dengan pendekatan eksploratif. Proses penelitian dilakukan melalui beberapa tahap sebagai berikut:

### 2.1. Desain Penelitian

Penelitian ini menggunakan metode *K-Means Clustering* untuk mengelompokkan lagu-lagu terpopuler di *Spotify* tahun 2024 berdasarkan karakteristik audio dan popularitasnya. Pendekatan ini bertujuan untuk mengidentifikasi pola dan tren dalam industri musik yang berkembang, serta memahami preferensi musik global.

### 2.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini diambil dari *dataset* "Most Streamed Spotify Songs 2024," yang tersedia di situs Kaggle. *Dataset* ini berformat CSV dan mencakup 4,600 baris data lagu dengan 29 atribut, termasuk nama lagu, nama album, nama artis, tanggal rilis, jumlah streaming, popularitas di berbagai *platform* musik (*Spotify*, *YouTube*, *TikTok*, *Apple Music*, *Deezer*, dll.), dan indikator konten eksplisit.

### 2.3. Prosedur Penelitian

Proses penelitian dimulai dengan preprocessing data, yang meliputi pembersihan data, penanganan nilai yang hilang, dan standarisasi data menggunakan *StandardScaler* dari pustaka *sklearn*. Data yang telah disiapkan kemudian dianalisis menggunakan metode *K-Means Clustering* untuk mengelompokkan lagu berdasarkan karakteristik audio dan metrik popularitasnya. Langkah-langkah utama dalam prosedur penelitian ini adalah sebagai berikut:

1. *Import Data*: *Dataset* diimpor menggunakan pustaka *Pandas*.
2. *Data Cleaning*: Data dibersihkan dari duplikasi dan *outlier*, serta nilai yang hilang diisi dengan nilai rata-rata atau median.
3. *Data Standardization*: Atribut numerik distandarisasi menggunakan *StandardScaler* untuk memastikan skala data yang konsisten.

4. *Clustering*: Algoritma K-Means diterapkan dengan jumlah *cluster* yang dioptimalkan menggunakan metode *Elbow* dan *Silhouette Score*. Setelah itu, *clustering* dilakukan untuk mengelompokkan lagu-lagu dalam *dataset*.

#### 2.4. Uji Validitas dan Evaluasi

Hasil dari *K-Means Clustering* dievaluasi dengan menggunakan *Silhouette Score* dan *Inertia* untuk menentukan kualitas klaster yang dihasilkan. Selain itu, centroid dari setiap *cluster* dianalisis untuk memahami karakteristik umum dari lagu-lagu dalam setiap *cluster*.

#### 2.5. Visualisasi Data

Visualisasi hasil *clustering* dilakukan dengan menggunakan pustaka Matplotlib dan Seaborn. Visualisasi ini mencakup *scatter plot* yang menampilkan distribusi lagu berdasarkan jumlah streaming di *Spotify* dan penayangan di YouTube, dengan setiap *cluster* diwakili oleh warna yang berbeda.

#### 2.6. Analisis dan Interpretasi

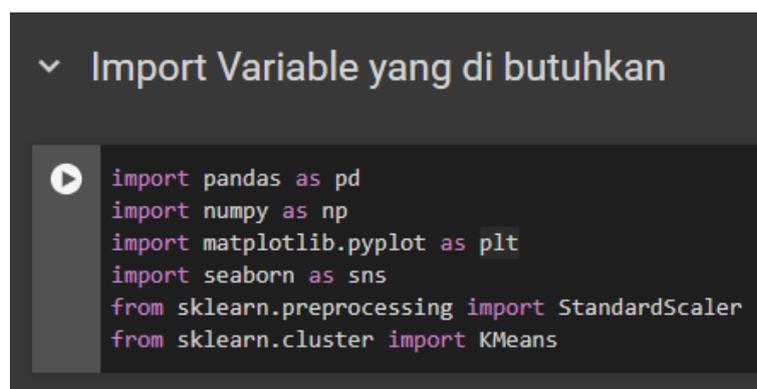
Setelah klasterisasi, dilakukan analisis lebih lanjut terhadap setiap *cluster* untuk mengidentifikasi pola-pola popularitas dan tren musik global yang diwakili oleh lagu-lagu dalam *cluster* tersebut. Hasil analisis ini digunakan untuk memberikan wawasan strategis bagi para pelaku industri musik.

Metode penelitian ini dirancang untuk memberikan pemahaman yang mendalam tentang bagaimana algoritma *K-Means Clustering* dapat digunakan untuk mengelompokkan lagu-lagu populer berdasarkan atribut audiens dan karakteristik popularitas mereka di berbagai *platform* streaming. Hasil penelitian ini diharapkan dapat memberikan kontribusi signifikan dalam literatur terkait analisis musik digital dan strategi pemasaran di industri musik.

### 3. Hasil dan Pembahasan

Pada penelitian ini, kami menggunakan *dataset* dari Website Kaggle.com yang berisikan kumpulan lagu dari aplikasi *Spotify*. *Dataset* ini berformat CSV dan mencakup 2001 data lagu. *Dataset* yang digunakan dalam penelitian ini, berjudul "Most Streamed *Spotify* Songs 2024," menawarkan wawasan komprehensif mengenai lagu-lagu yang paling banyak didengarkan di *Spotify* pada tahun 2024. Data ini mencakup atribut-atribut lagu, popularitas, dan kehadiran di berbagai *platform* musik, menjadikannya sumber yang berharga bagi analis musik, penggemar, dan profesional industri musik.

Beberapa fitur kunci dari *dataset* ini meliputi nama lagu, nama album, nama artis, tanggal rilis, ISRC (*International Standard Recording Code*), peringkat sepanjang masa, skor lagu, jumlah streaming di *Spotify*, jumlah dan jangkauan *playlist Spotify*, popularitas di *Spotify*, jumlah tampilan di YouTube, jumlah suka di YouTube, jumlah postingan TikTok yang menampilkan lagu tersebut, jumlah suka di TikTok, jumlah tampilan di TikTok, jangkauan *playlist* di YouTube, jumlah *playlist* di *Apple Music*, jumlah pemutaran di radio (*AirPlay Spins*), jumlah pemutaran di SiriusXM, jumlah *playlist* di *Deezer*, jangkauan *playlist* di *Deezer*, jumlah *playlist* di *Amazon Music*, jumlah streaming di Pandora, jumlah stasiun di Pandora yang menampilkan lagu tersebut, jumlah streaming di *Soundcloud*, jumlah *Shazam*, popularitas di TIDAL, dan indikasi apakah lagu tersebut mengandung konten eksplisit.



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

Gambar 1 *Import Variable*

Dalam analisis ini, kami menggunakan pustaka *pandas* untuk manipulasi data, *numpy* untuk komputasi numerik, dan *matplotlib.pyplot* serta *seaborn* untuk visualisasi data. *Standard Scaler* dari *sklearn* digunakan untuk menstandarisasi data, dan *KMeans* dari *sklearn* digunakan untuk mengelompokkan data, membantu mengidentifikasi pola dalam *dataset*.

```
# Load the data from the uploaded CSV file
df = pd.read_csv('/content/Most Streamed Spotify Songs 2024.csv', encoding='latin1')
print(df)
```

	Track	Album Name
0	MILLION DOLLAR BABY	Million Dollar Baby - Single
1	Not Like Us	Not Like Us
2	i like the way you kiss me	I like the way you kiss me
3	Flowers	Flowers - Single
4	Houdini	Houdini
...	...	...
4595	For the Last Time	For the Last Time
4596	Dil Meri Na Sune	Dil Meri Na Sune (From "genius")
4597	Grace (feat. 42 Dugg)	My Turn
4598	Nashe Si Chadh Gayi	November Top 10 Songs
4599	Me Acostumbre (feat. Bad Bunny)	Me Acostumbre (feat. Bad Bunny)

	Artist	Release Date	ISRC	All Time Rank	Track Score
0	Tommy Richman	4/26/2024	QM24S2402528	1	725.4
1	Kendrick Lamar	5/4/2024	USUG12400910	2	545.9
2	Artemas	3/19/2024	QZ3842400387	3	538.4
3	Miley Cyrus	1/12/2023	USSM12209777	4	444.9
4	Eminem	5/31/2024	USUG12403398	5	423.3
...	...	...	...	...	...
4595	\$uicideboy\$	9/5/2017	QM8DG1703420	4,585	19.4
4596	Atif Aslam	7/27/2018	INT101800122	4,575	19.4
4597	42 Dugg	7/28/2020	USUG12000042	4,571	19.4

Gambar 3.2 Mengolah Data

Dataset yang digunakan dalam penelitian ini berisikan data lagu-lagu yang paling banyak didengarkan di Spotify pada tahun 2024. Dataset ini terdiri dari 4600 baris dan 29 kolom, mencakup informasi seperti judul lagu, nama album, nama artis, tanggal rilis, kode ISRC, peringkat sepanjang masa, skor lagu, jumlah streaming di Spotify, popularitas di berbagai platform musik seperti YouTube, TikTok, Apple Music, Deezer, Amazon Music, Pandora, dan Soundcloud, serta beberapa metrik lainnya seperti jumlah Shazam dan apakah lagu tersebut mengandung konten eksplisit. Data ini menyediakan wawasan komprehensif mengenai tren musik dan popularitas lagu-lagu di berbagai platform streaming.

```
Number of data points in each cluster:
Cluster
0 3357
3 918
1 192
2 131
4 2
Name: count, dtype: int64

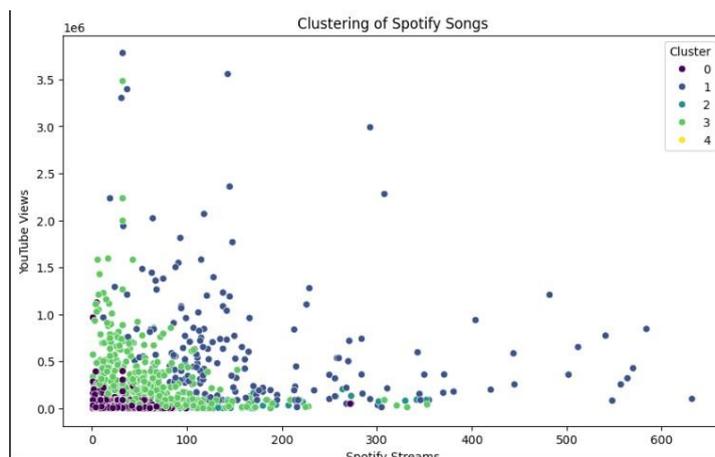
Centroids of each cluster:
AirPlay Spins Amazon Playlist Count Apple Music Playlist Count \
0 29436.301683 19.305956 29.039812
1 185010.722016 63.307536 232.061507
2 390538.307737 92.918694 174.882467
3 74272.269433 29.918579 93.980643
4 1386.000000 6.500000 6.500000

Deezer Playlist Count Deezer Playlist Reach Pandora Streams \
0 18.248144 1.056165e+06 4.091834e+07
1 170.640275 1.528471e+06 4.549846e+08
2 100.732824 1.081256e+07 6.693646e+07
3 45.126704 7.633050e+05 1.749777e+08
4 6.500000 1.272290e+05 4.288739e+07

Pandora Track Stations Shazam Counts SiriusXM Spins Soundcloud Streams \
0 38751.961446 1.075316e+06 229.164238 1.319979e+07
1 626500.627179 1.375143e+07 308.560120 2.033657e+07
```

Gambar 3.3 Perhitungan Clustering

Analisis clustering menunjukkan bahwa dalam dataset, Cluster 0 terdiri dari 3,357 lagu yang memiliki jumlah putaran yang moderat. Cluster 1, dengan 192 lagu, berisi lagu-lagu yang sangat populer dan memiliki jumlah putaran yang sangat tinggi. Cluster 2, yang berisi 131 lagu, juga mencakup lagu-lagu populer, namun dengan variasi dalam jumlah putaran. Cluster 3 terdiri dari 918 lagu dengan popularitas menengah. Cluster 4 sangat unik dengan hanya 2 lagu yang memiliki jumlah putaran sangat rendah. Centroid dari setiap cluster memberikan gambaran tentang rata-rata fitur dari lagu-lagu dalam cluster tersebut. Sebagai contoh, Cluster 0 memiliki rata-rata AirPlay Spins sebesar 29,436, Pandora Streams sebesar 40,918,340, dan Spotify Streams sebesar 207,008,500. Cluster 1 memiliki rata-rata AirPlay Spins sebesar 185,011, Pandora Streams sebesar 454,984,600, dan Spotify Streams sebesar 1,806,797,000, menunjukkan bahwa lagu-lagu dalam cluster ini sangat populer. Sementara itu, Cluster 2 dan Cluster 3 memiliki jumlah putaran yang signifikan namun bervariasi, dan Cluster 4 menonjol dengan rata-rata yang sangat rendah dalam semua fitur, mencerminkan bahwa lagu-lagu dalam cluster ini sangat unik dengan jumlah putaran yang sangat rendah. Analisis ini membantu mengidentifikasi pola popularitas lagu di berbagai platform streaming utama seperti Spotify dan YouTube.



Gambar 3.4 Hasil Visualisasi

Hasil visualisasi *clustering* menunjukkan distribusi lagu berdasarkan jumlah putaran di *Spotify* (sumbu X) dan jumlah penayangan di YouTube (sumbu Y). *Cluster 0*, yang diwakili oleh warna ungu, mendominasi dengan 3,357 lagu yang memiliki jumlah putaran *Spotify* dan penayangan YouTube yang rendah hingga sedang. *Cluster 1* (biru) terdiri dari 192 lagu yang sangat populer di YouTube, dengan jumlah penayangan yang tinggi tetapi variasi dalam jumlah putaran *Spotify*. *Cluster 2* (hijau) memiliki 131 lagu yang menunjukkan distribusi sedang hingga tinggi baik pada putaran *Spotify* maupun penayangan YouTube. *Cluster 3* (kuning) mencakup 918 lagu dengan popularitas menengah di kedua *platform*. Terakhir, *Cluster 4*, yang sangat kecil dengan hanya 2 lagu, tidak memberikan informasi yang signifikan. Secara keseluruhan, visualisasi ini membantu mengidentifikasi pola popularitas lagu di dua *platform* streaming utama, *Spotify* dan YouTube.

#### 4. Kesimpulan

Penelitian ini memberikan wawasan mendalam tentang tren musik dan pola popularitas lagu di *Spotify* tahun 2024. Melalui metode *K-Means Clustering*, lima *cluster* utama berhasil diidentifikasi, masing-masing dengan karakteristik unik yang mencerminkan preferensi musik global. Hasil analisis menunjukkan bahwa lagu-lagu populer memiliki variasi dalam jumlah putaran di berbagai *platform* streaming, seperti *Spotify* dan YouTube, dengan beberapa lagu menunjukkan dominasi yang signifikan di *cluster* tertentu. Temuan ini memberikan kontribusi penting bagi literatur musik digital dan industri musik, membantu para pelaku industri dalam mengembangkan strategi promosi dan distribusi yang lebih efektif. Selain itu, hasil penelitian ini juga membuka peluang untuk studi lanjutan mengenai pengaruh *platform* streaming terhadap industri musik secara keseluruhan, termasuk dampak sosial dan budaya serta implikasi ekonomi dari perubahan dalam industri musik. Dengan memahami pola popularitas lagu, para artis dan produser dapat menciptakan karya yang lebih sesuai dengan preferensi pasar, sementara akademisi dan peneliti dapat menjadikan temuan ini sebagai dasar untuk eksplorasi lebih lanjut dalam bidang musik digital dan analisis data.

#### Daftar Pustaka

- [1] J. Doe, "Housing Prices in California," Kaggle, 2022. [Online]. Available: <https://www.kaggle.com/johndoe/housing-prices-california>. [Accessed: 25 Jul. 2024].
- [2] B. Winarso, "Berkenal dengan Layanan Streaming Musik, *Spotify*," Hybrid.co.id, Apr. 6, 2016. [Online]. Available: URL. [Accessed: 25 Jul. 2024].
- [3] R. C. Maringka, A. Khoirunnita, R. Maringka, and E. Utami, "Analisa Perkembangan Musik Pada *Spotify* Menggunakan Structured Query Language (SQL)," CogITO Smart Journal, vol. 7, no. 1, pp. 1-14, Jun. 2021, doi: 10.31154/cogito.v7i1.287.
- [4] R. Fiska, "Pengertian seni musik: Sejarah, unsur, jenis dan fungsinya," Gramedia Literasi, n.d. [Online]. Available: <https://www.gramedia.com/literasi/seni-musik/>. [Accessed: 25 Jul. 2024].
- [5] *Spotify*, "Apa itu *Spotify*?," Situs Dukungan *Spotify*, 2024. [Online]. Available: <https://support.spotify.com/id-id/article/what-is-spotify>. [Accessed: 25 Jul. 2024].
- [6] Universitas Stekom, "Psikologi Musik," Pusat Ensiklopedia, 2024. [Online]. Available: [https://p2k.stekom.ac.id/ensiklopedia/Psikologi\\_musik](https://p2k.stekom.ac.id/ensiklopedia/Psikologi_musik). [Accessed: 25 Jul. 2024].

- 
- [7] R. Setiawan, "Apa itu Data Mining dan Bagaimana Metodenya?," Dicoding Indonesia, 30 Oct. 2021. [Online]. Available: <https://www.dicoding.com/blog/apa-itu-data-mining/>. [Accessed: 25 Jul. 2024].
- [8] RevoU, "K Means *Clustering*," Revoupedia, 2024. [Online]. Available: <https://revou.co/k-means-clustering>. [Accessed: 28 Jul. 2024].afD