

Clustering Korean Drama Viewers' Preferences for Marketing Strategy Optimization Using the K-Means Algorithm on the MyDramaList Platform

Sulistina Nabilla ^{1*}, Agung Nugroho ^{2*}, Isria Miharti Mahaerni Putri ^{3*}

* Teknik Informatika, Universitas Pelita Bangsa

sulistinabilla@gmail.com ¹, agung@pelitabangsa.ac.id ², isriamiharti@pelitabangsa.ac.id ³

Article Info

Article history:

Received 2025-06-15

Revised 2025-06-28

Accepted 2025-07-03

Keyword:

*Korean Drama,
K-Means Clustering,
Audience Preferences,
Marketing Strategy,
MyDramaList.*

ABSTRACT

Korean dramas are highly popular in Indonesia, but it is a challenge for marketers to understand the diverse tastes of the audience. This study aims to identify audience preference segments by applying the K-Means algorithm to perform clustering. The analysis was conducted on 500 Korean dramas from the MyDramaList platform released in the period 2020 to 2023. Features used in the clustering process include rating, number of viewers (*no_of_viewers*), year of release (*year*), and genre. To overcome the multi-label genre, this research uses one-hot encoding technique to convert categorical data into numerical format. The optimal number of clusters was determined as four (4) based on analysis using the Elbow Method. The analysis successfully identified four distinct audience segments. Cluster 1 is the largest market segment which includes 323 dramas with a dominant genre of "Drama and Romance" and an average rating of 7.64. In contrast, Cluster 2 is a high-quality niche segment consisting of only 16 dramas but has the highest average rating (8.29) as well as the highest average viewership (25,739), with the dominant genre of "Drama and Mystery". The other two segments are Cluster 0 (58 dramas) which focuses on the "Thriller" and "Mystery" genres, and Cluster 3 (147 dramas) which features the "Comedy, fantasy and romance" genres. These data-driven findings enable the development of more specific and targeted marketing strategies for each audience profile, thereby improving promotional effectiveness as well as the viewing experience.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Industri hiburan global telah menunjukkan lonjakan pertumbuhan yang luar biasa dalam beberapa dekade terakhir, menjadikannya salah satu bidang hiburan paling berpengaruh di dunia. Kemajuan ini terlihat dari semakin banyaknya produksi drama berkualitas yang tidak hanya memenuhi kebutuhan penonton domestik tetapi juga berhasil menarik perhatian pasar internasional. Keberhasilan ini tidak lepas dari daya tarik cerita yang inovatif, kualitas produksi yang tinggi, dan strategi pemasaran yang cermat. Drama Korea (K-Drama) kini telah menjadi bagian yang tidak terpisahkan dari budaya populer global, terutama di kawasan Asia, termasuk Indonesia, dengan pengaruh yang kian meluas [1].

Di Indonesia, K-Drama berhasil membangun basis penggemar yang besar dan aktif, yang terdiri dari berbagai kelompok usia serta latar belakang sosial. Hal ini mencerminkan popularitasnya yang terus meningkat, ditandai dengan partisipasi aktif dalam diskusi daring, rekomendasi drama kepada sesama penggemar, hingga keterlibatan dalam komunitas pecinta Korea [2]. Fenomena ini mengindikasikan bagaimana tema-tema dalam K-Drama, yang sering kali mencerminkan kehidupan sehari-hari, berhasil menyentuh sisi emosional dan psikologis penontonnya. Berdasarkan data yang dirilis oleh The Korea Times pada tahun 2018, terjadi peningkatan signifikan dalam jumlah penggemar kebudayaan Korea di seluruh dunia. Jumlah tersebut melonjak sebesar 22 persen, mencapai angka 89,19 juta penggemar dibandingkan dengan

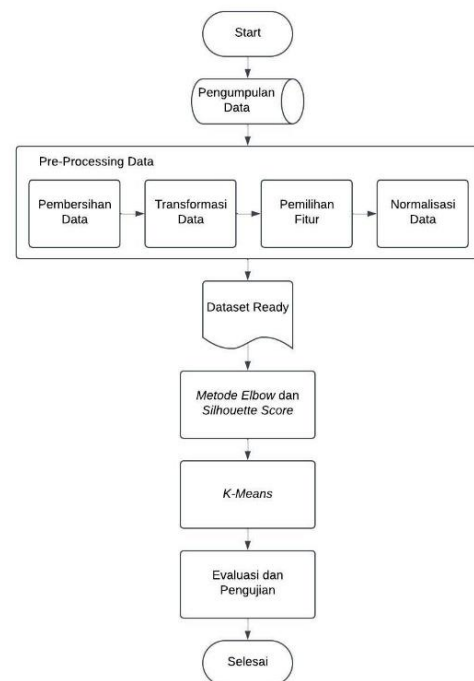
tahun sebelumnya, yang hanya mencatatkan 73,12 juta penggemar pada 2017 [3]. Tren ini menunjukkan semakin luasnya pengaruh budaya Korea secara global, mencakup berbagai aspek seperti K-Pop, drama, makanan, dan tradisi yang semakin digemari di berbagai belahan dunia. Fenomena Korean Wave (Hallyu) telah berkembang dari sekadar tren budaya menjadi salah satu alat diplomasi budaya yang signifikan. Penyebaran budaya Korea melalui media hiburan seperti drama dan musik tidak hanya memperkuat citra budaya Korea di mata dunia tetapi juga meningkatkan daya saing ekonomi Korea Selatan di pasar global. Dampak ini terlihat dari meningkatnya minat terhadap berbagai produk Korea, mulai dari kosmetik hingga makanan, yang memberikan kontribusi besar terhadap pertumbuhan ekonomi dan memperluas jangkauan industri kreatif Korea Selatan [4].

Dalam hal perilaku audiens, platform daring seperti MyDramaList memiliki peranan penting sebagai sumber informasi dan ruang diskusi bagi penggemar K-Drama. Fitur seperti ulasan, *rating*, dan rekomendasi memungkinkan pengguna untuk saling berbagi pengalaman menonton serta memengaruhi keputusan dalam memilih drama yang diminati. Data yang dihasilkan dari ulasan dan interaksi pengguna dapat digunakan untuk memahami pola preferensi penonton secara lebih mendalam dan akurat [5]. Guna memahami lebih dalam faktor-faktor yang memengaruhi preferensi penonton, diperlukan pendekatan analisis yang sistematis dan berbasis data. Analisis terhadap preferensi audiens menjadi alat yang sangat penting untuk membantu produsen konten maupun pemasar dalam menyusun strategi yang lebih relevan. Penggunaan teknik *clustering* dapat memberikan wawasan yang lebih mendalam tentang perilaku penonton, sehingga membantu dalam pengambilan keputusan strategis [6]. Metode *clustering* berdasarkan genre menawarkan pendekatan inovatif untuk memahami preferensi audiens dengan lebih terperinci. Teknik ini memungkinkan pemetaan kategori drama berdasarkan genre dan karakteristik lainnya, sehingga pemasar dapat menciptakan kampanye yang lebih terfokus dan efektif. Studi terdahulu menunjukkan bahwa penerapan metode *clustering* dapat meningkatkan efisiensi strategi pemasaran dengan memanfaatkan data perilaku konsumen yang tersedia [7]. Di antara berbagai algoritma yang digunakan dalam analisis *clustering*, **K-Means** menonjol karena efisiensi dan kemampuannya untuk menghasilkan pengelompokan data yang akurat dalam waktu singkat. Algoritma ini dianggap ideal untuk menganalisis data dengan jumlah besar karena kesederhanaannya dalam implementasi dan interpretasi. Penggunaan K-Means dapat menghasilkan pengelompokan yang mudah dipahami dan dapat diaplikasikan pada berbagai skenario analisis data [6]. Beberapa penelitian sebelumnya menunjukkan bagaimana algoritma K-Means telah digunakan secara efektif dalam berbagai konteks untuk mengidentifikasi pola dan tren dalam data. Sebagai contoh, peneliti terdahulu menunjukkan penerapan algoritma ini

dalam menganalisis perilaku konsumen, yang hasilnya dapat diterapkan untuk merancang strategi pemasaran yang lebih tepat sasaran [8]. Hubungan antara genre drama dan preferensi penonton adalah salah satu aspek penting yang perlu dikaji lebih dalam. Genre bukan sekadar kategori hiburan, melainkan juga cerminan dari kebutuhan emosional dan psikologis audiens. Teknik pengelompokan dapat mengungkap pola-pola tertentu yang tidak terlihat secara kasat mata, membantu memahami keterkaitan antara minat audiens dengan genre drama favorit mereka [9].

II. METODE PENELITIAN

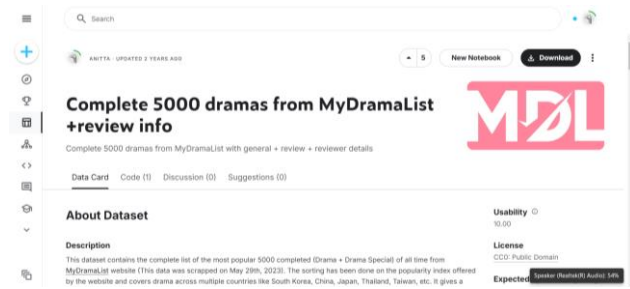
Metode penelitian ini terdiri dari beberapa tahapan yang dirancang untuk mengelompokkan preferensi penonton drama Korea berdasarkan atribut seperti genre dan rating menggunakan algoritma K-Means. Maka dari itu, penulis membuat alur atau tahapan dalam penelitian ini lebih jelas dapat dilihat pada gambar 1.



Gambar 1. Metode Penelitian

A. Pengumpulan Data

Teknik pengumpulan data dilakukan dengan mengunduh dataset dari Kaggle dengan judul "Complete 5000 Dramas from MyDramaList - Review Info" pada platform MyDramaList". Tampilan sumber data lebih jelas dapat dilihat pada gambar 2.



Gambar 2. Sumber Data

Gambar 3. Data Awal

Pada gambar 3 Dataset tersebut mencakup data sebanyak 5000 mengenai drama dari berbagai negara di seluruh dunia, yang dirancang untuk memberikan informasi yang komprehensif terkait berbagai aspek drama tersebut. Drama dalam dataset ini mencakup periode waktu dari tahun 1971 hingga 2023. Dataset ini memiliki total 33 atribut, yang mencakup berbagai elemen penting, atribut pada dataset awal lebih jelas dapat dilihat pada tabel 1.

TABEL 1
ATRIBUT PADA DATASET AWAL

Atribut	Keterangan
<i>Unnamed: 0</i>	Kolom kosong untuk urutan drama
<i>name</i>	Judul drama
<i>content</i>	Deskripsi singkat atau sinopsis dari drama
<i>no_of_reviews</i>	Jumlah ulasan yang diterima oleh drama tersebut.
<i>no_of_viewers</i>	Jumlah penonton atau orang yang telah melihat drama tersebut.
<i>aka_names</i>	Nama lain atau judul alternatif drama.
<i>screenwriter</i>	Nama penulis naskah dari drama.
<i>director</i>	Nama sutradara dari drama.
<i>genres</i>	Genre atau kategori dari drama (misalnya: drama, komedi, aksi).
<i>Tags</i>	Kata kunci yang menggambarkan drama, biasanya terkait tema atau elemen cerita.
<i>country</i>	Negara asal produksi drama.
<i>type_of_show</i>	Jenis acara, misalnya serial TV, film, dokumenter, atau lainnya.
<i>episodes</i>	Jumlah episode dalam
<i>end_date</i>	Tanggal berakhirnya penayangan drama.

<i>start_date</i>	Tanggal mulai penayangan drama.
<i>year</i>	Tahun rilis acara/film.
<i>duration</i>	Durasi drama (jam/menit/detik).
<i>rating</i>	Nilai atau skor yang diberikan untuk drama.
<i>no_of_rating</i>	Jumlah pengguna yang memberikan rating.
<i>rank</i>	Peringkat drama berdasarkan popularitas atau skor.
<i>popularity</i>	Tingkat popularitas drama, sering diukur dari jumlah penonton, ulasan, atau rating.
<i>content_rating</i>	Klasifikasi usia atau panduan penonton (misalnya: PG-13, R, Semua Umur).
<i>where_to_watch</i>	Platform atau layanan streaming tempat drama dapat ditonton.
<i>main_role</i>	Nama aktor/aktris utama yang membintangi drama.
<i>support_role</i>	Nama aktor/aktris pendukung dalam acara/film.
<i>no_of_extracted_reviews</i>	Jumlah ulasan yang berhasil diambil untuk analisis.
<i>Total_sentences</i>	Total jumlah kalimat yang dianalisis dari ulasan.
<i>POSITIVE_people_sentiment</i>	Persentase atau jumlah orang yang memiliki sentimen positif terhadap drama.
<i>POSITIVE_sentences</i>	Jumlah kalimat yang mengandung sentimen positif dalam ulasan.
<i>NEGATIVE_people_sentiment</i>	Persentase atau jumlah orang yang memiliki sentimen negatif terhadap acara/film.
<i>NEGATIVE_sentences</i>	Jumlah kalimat yang mengandung sentimen negatif dalam ulasan.
<i>reviewer_location_info</i>	Informasi lokasi para pengulas.
<i>reviewer_gender_info</i>	Informasi jenis kelamin pengulas.

Pada gambar 3 dataset tersebut kemudian diolah secara manual dengan memberikan perhatian khusus pada data yang berkaitan dengan drama Korea. Proses pengolahan ini difokuskan pada drama yang dirilis dalam rentang waktu tahun 2020 hingga 2023, dengan tujuan untuk menyaring dan menyesuaikan data agar relevan dengan analisis yang dilakukan. Pemilihan periode ini dilakukan karena dianggap mewakili perkembangan signifikan dalam industri drama Korea, baik dari segi popularitas global maupun kualitas produksi.

	A	B	C	D	E
1	name	no_of_viewers	genres	year	rating
2	it's okay to not be okay	156581	Psychological, Comedy, Romance, Dr	2020	8,9
3	itawon class	116763	Business, Romance, Life, Drama	2020	8,5
4	true beauty	109808	Comedy, Romance, Youth, Drama	2020	8,3
5	flower of evil	99513	Thriller, Romance, Crime, Melodram	2020	9,1
6	big picture house	908	Music, Romance, Youth, Drama	2020	7,3
7	handmade love	6160	Romance, Fantasy	2020	7,9
8	more than friends	19900	Comedy, Romance, Life, Drama	2020	7,3
9	no going back romance	1443	Mystery, Romance, Youth	2020	7,3
10	discipline z: vampire	4175	Mystery, Romance, Sci-Fi, Supernatu	2020	6,3
11	flower of evil special	1281	Thriller, Documentary	2020	8,1
12	sf8: the prayer	1280	Drama, Sci-Fi	2020	7,6
13	touch me if you can	2544	Romance, Fantasy	2020	7
14	café kilimanjaro	1020	Music, Drama	2020	7,3
15	sf8: white crow	1018	Horror, Psychological, Sci-Fi	2020	7,5
16	pop out boy!	7944	Comedy, Romance, Drama, Fantasy	2020	7,7
17	born again	14285	Mystery, Romance, Fantasy, Melodra	2020	7,3
18	the king: eternal monarch	88121	Mystery, Romance, Drama, Sci-Fi	2020	8,1
19	hospital playlist	84596	Romance, Life, Drama, Medical	2020	9,1
20	start-up	84508	Business, Comedy, Romance, Youth	2020	8,1
21	tale of the nine-tailed	81768	Thriller, Horror, Romance, Fantasy	2020	8,3
22	sweet home	79202	Action, Horror, Drama, Sci-Fi	2020	8,7

Gambar 4. Data yang digunakan

Setelah melalui proses seleksi dan penyaringan, Pada gambar 4 dataset akhir atau data yang digunakan terdiri dari 500 data yang fokus pada drama Korea. Dataset ini mencakup 5 atribut utama yang dianggap relevan untuk analisis lebih mendalam. Atribut-atribut tersebut dipilih secara cermat untuk memastikan representasi yang akurat terhadap berbagai aspek drama, atribut pada dataset yang digunakan lebih jelas dapat dilihat pada tabel 1.

TABEL 2
ATRIBUT PADA DATASET YANG DIGUNAKAN

Atribut	Keterangan
name	Judul drama
no_of_viewers	Jumlah penonton atau orang yang telah melihat drama tersebut.
genres	Genre atau kategori dari drama (misalnya: drama, komedi, aksi).
year	Tahun rilis acara/film.
rating	Nilai atau skor yang diberikan untuk drama.

B. Pre-Processing Data

Dalam tahap pre-processing/pembersihan, proses akan dilakukan untuk menghasilkan dataset yang bersih sehingga dapat digunakan dalam tahap mining untuk mengumpulkan informasi bermanfaat [10]. Proses ini mencakup beberapa langkah berikut:

1) Pembersihan Data

Cleaning data, juga dikenal sebagai pembersihan data, adalah proses mengidentifikasi, mengatasi, dan memperbaiki masalah data atau ketidakakuratan dalam kumpulan data. Tujuan utama dari pembersihan data adalah untuk memastikan data bersih, konsisten, dan valid untuk digunakan dalam analisis atau pemodelan. dilakukan untuk menghilangkan noise, outlier, dan nilai yang tidak ada dalam data. Noise dan outlier dapat menyebabkan analisis data yang

tidak akurat, sedangkan nilai yang tidak ada dapat mengganggu proses analisis data [11].

2) Transformasi Data

Setelah data dibersihkan, tahap berikutnya adalah Proses mengubah data menjadi format yang dibutuhkan agar dapat diproses pada tahap data mining dikenal sebagai transformasi data [12]. Salah satu perubahan yang dilakukan adalah mengonversi kolom genres yang berisi teks menjadi angka dengan menggunakan multi-hot encoding. Metode ini mengubah setiap genre menjadi kolom biner, di mana angka 1 menunjukkan genre tersebut ada, dan angka 0 menunjukkan genre tersebut tidak ada. Setelah itu, hasil encoding digabungkan kembali ke dalam dataset agar informasi genre tetap ada dalam bentuk numerik.

3) Pemilihan Fitur

Metode preprocessing penting untuk klasifikasi adalah pemilihan fitur, yang dapat mengurangi kinerja klasifikasi [13]. Pemilihan subset darigifiturgasli sesuai dengan kriteria seleksi fitur dikenal sebagai seleksi fitur. Hasil seleksi fitur termasuk mengurangi waktu komputasi, meningkatkan akurasi pembelajaran, dan menyederhanakan hasil pembelajaran [13]. Pada tahap ini, dipilih fitur-fitur yang penting untuk proses analisis, seperti clustering. Fitur yang dipilih bisa mencakup kolom genre, durasi, rating, atau tahun rilis yang dianggap relevan untuk analisis. Memilih fitur yang tepat sangat penting agar model dapat menemukan pola yang bermakna dalam data.

4) Normalisasi Data

Normalisasi dilakukan untuk memastikan bahwa semua fitur berada pada skala yang sama. Ini penting karena beberapa algoritma, seperti clustering, sensitif terhadap perbedaan skala antar fitur. Dengan normalisasi, fitur dengan skala besar tidak akan mendominasi hasil analisis. Normalisasi dilakukan dengan metode seperti Min-Max scaling atau Z-score standardization untuk menyesuaikan skala data sebelum digunakan dalam model.

C. Penerapan Metode Elbow dan Silhouette Score

Metode elbow adalah teknik yang menentukan jumlah cluster optimal dengan mengidentifikasi titik di grafik di mana terjadi penurunan signifikan, membentuk lekukan tajam. Titik ini menunjukkan nilai k yang paling sesuai. Nilai k optimal dapat ditemukan dengan menganalisis grafik WCSS (Within Cluster Sum of Squares) atau biasa disebut dengan Sum of Square Error (SSE) dan membandingkan nilainya pada berbagai jumlah kluster [14].

Metode elbow bertujuan untuk memilih nilai k terkecil dengan nilai internal yang rendah. Jumlah cluster optimal ditentukan dengan membandingkan nilai SSE pada setiap jumlah cluster. Seiring bertambahnya jumlah cluster, grafik akan membentuk pola seperti siku; semakin besar nilai k, semakin kecil nilai SSE [15].

Silhouette Score adalah metrik yang digunakan untuk mengevaluasi kualitas clustering data dengan mengukur jarak antara data point dalam clusternya sendiri dan jarak antara data point dengan cluster terdekat lainnya. Nilai Silhouette yang lebih tinggi menunjukkan bahwa data point diklasifikasikan lebih baik dalam cluster mereka. Nilai Silhouette yang berkisar antara -1 dan 1 menunjukkan bahwa data point diklasifikasikan lebih baik dalam cluster mereka [16]. Dengan demikian, semakin tinggi nilai Silhouette Score, maka kualitas pemisahan cluster yang terbentuk semakin baik dan representatif.

Dalam penelitian ini, metode Elbow dan Silhouette Score diterapkan untuk menentukan klaster optimal pada data yang dianalisis. Hasilnya digunakan untuk memastikan segmentasi data yang efisien dan akurat.

D. Penerapan K-Means

K-Means adalah algoritma clustering yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok (cluster) berdasarkan kesamaan fitur. Tujuan dari algoritma ini adalah untuk meminimalkan jarak antara data dan pusat cluster (centroid), dengan cara membagi data ke dalam K cluster yang telah ditentukan sebelumnya. K-Means dipilih dalam penelitian ini karena kesederhanaannya, efisiensinya dalam mengolah data besar, serta kemampuannya untuk mengidentifikasi pola dalam data yang memiliki cluster yang jelas dan terpisah.

Langkah-langkah dalam penggunaan algoritma K-Means Clustering adalah sebagai berikut [17]:

- Memilih jumlah cluster (K) yang akan dibentuk.
- Inisialisasi nilai centroid awal untuk masing masing cluster.
- Gunakan metode Euclidean Distance yang ada di bawah ini untuk menghitung jarak antara setiap objek dengan centroid yang sesuai. Setelah dihitung, setiap objek dimasukkan ke dalam centroid dengan nilai centroid terdekat/minimum.

$$d_{ik} = \sqrt{\sum_{j=1}^m (X_{ij} - C_{kj})^2}$$

d_{ik} : Jarak antara data ke-i dengan centroid tiap cluster

m : jumlah atribut

X_{ij} : data ke-i

C_{kj} : data pusat tiap cluster.

- Setiap objek pada cluster tersebut akan dihitung menjadi nilai centroid baru.
- Ulangi langkah 3 hingga posisi centroid sudah tidak mengalami perubahan.

E. Evaluasi dan Pengujian

Evaluasi dan pengujian dalam penelitian ini dilakukan untuk menilai seberapa baik algoritma K-Means dalam

mengelompokkan data sesuai pola pada dataset dan tujuan penelitian. Evaluasi dilakukan untuk memastikan jumlah cluster (K) yang ditentukan tepat serta hasil clustering mencerminkan struktur data yang sebenarnya, dengan menggunakan metrik seperti Inertia untuk mengukur kualitas pengelompokan. Sementara itu, pengujian bertujuan mengevaluasi akurasi algoritma dalam menghasilkan pola yang konsisten dan dapat diandalkan pada data baru (data testing). Kedua proses ini membantu peneliti memahami kinerja model secara objektif serta menilai kelebihan dan kekurangannya, sehingga hasil yang diperoleh tidak hanya mencerminkan kebetulan tetapi juga pola yang sesungguhnya.

III. HASIL DAN PEMBAHASAN

A. Deskripsi Dataset

Penelitian ini dilakukan untuk mendapatkan optimalisasi strategi untuk pemasaran dari clustering genre yang di harapkan bisa menjadi membantu pemasaran agar lebih tepat dalam meningkatkan MyDramalist agar bisa menjadi situs film yang diminati oleh penggemar drama korea. Langkah yang dilakukan penulis adalah deksripsi data set dan data yang digunakan berasal dari situs web Kaggle dengan judul "Complete 5000 Dramas from MyDramaList - Review Info" pada platform MyDramaList yang merupakan salah satu platform terkemuka dalam menyediakan informasi, rating, serta ulasan terkait drama Korea dan Asia. Dataset yang berhasil dikumpulkan terdiri dari 500 data drama Korea, dengan 4 atribut utama yang akan digunakan dalam proses pengolahan data. Dataset ini digunakan untuk mengelompokkan drama berdasarkan kesamaan karakteristik menggunakan metode K-Means Clustering, sehingga dapat diketahui segmen-segmen pasar drama Korea yang berbeda berdasarkan data numerik dan kategorikal.

TABEL 3
Atribut dataset

Atribut	Keterangan
<i>Rating</i>	Nilai atau skor yang diberikan untuk drama.
<i>genres</i>	Genre atau kategori dari drama (misalnya: drama, komedi, aksi).
<i>year</i>	Tahun rilis film
<i>no_of_viewers</i>	Jumlah penonton atau orang yang telah melihat drama tersebut.

Data ini kemudian diproses lebih lanjut melalui serangkaian tahapan pra-pemrosesan guna mempersiapkannya untuk digunakan dalam proses klasterisasi. Proses ini meliputi pembersihan data (data cleaning) untuk mengatasi nilai yang hilang atau tidak konsisten, transformasi data untuk mengubah tipe data yang bersifat kategorikal ke bentuk numerik, serta normalisasi

agar skala antar atribut menjadi seragam. Tujuan dari tahapan ini adalah untuk memastikan bahwa data yang digunakan dalam algoritma K-Means berada dalam kondisi yang optimal, sehingga hasil pengelompokan yang dihasilkan menjadi lebih akurat dan dapat diinterpretasikan dengan baik. Dengan kualitas data yang telah disiapkan tersebut, maka proses klasterisasi dapat dilakukan untuk mengidentifikasi pola tersembunyi serta segmentasi karakteristik drama Korea yang dianalisis dalam penelitian ini.

B. Praproses Data

Pra-pemrosesan data merupakan tahap krusial sebelum dilakukan analisis dengan algoritma K-Means. Tujuan dari proses ini adalah untuk memastikan bahwa data yang digunakan bersih, relevan, dan berada dalam bentuk yang dapat dianalisis secara efektif. Proses pra-pemrosesan dilakukan dengan beberapa tahapan sebagai berikut:

1) Pemuatan Dataset

Dataset yang digunakan merupakan data sekunder hasil scraping dari situs MyDramaList, yang berisi informasi 5.000 drama populer dari berbagai negara. Dataset ini dimuat menggunakan pustaka pandas dengan fungsi:

```
df=pd.read_csv("Top_5000_popular_drama_details_from_mydramalist.csv")
```

2) Filterisasi Data Korea Selatan (2020-2023)

Untuk memastikan fokus penelitian hanya pada drama Korea, dilakukan penyaringan dengan kondisi:

- Negara produksi adalah South Korea
- Tahun penayangan antara 2020 hingga 2023

```
df_korea = df[(df['country'].str.contains('South Korea', case=False, na=False)) & (df['year'] >= 2020) & (df['year'] <= 2025)]
```

3) Penanganan Nilai Kosong

Nilai kosong dapat menyebabkan gangguan dalam proses clustering. Oleh karena itu, dilakukan penanganan sebagai berikut:

- Kolom `genres` diisi dengan label "Unknown" jika kosong.
- Kolom `rating` dan `no_of_viewers` diisi dengan angka nol (0) untuk menjaga konsistensi numerik.

```
df_korea.loc[:, 'genres'] = df_korea['genres'].fillna('Unknown')
df_korea.loc[:, 'rating'] = df_korea['rating'].fillna(0)
df_korea.loc[:, 'no_of_viewers'] = df_korea['no_of_viewers'].fillna(0)
```

4) Encoding Genre (One-Hot Encoding)

Karena `genres` berupa teks multikategori (contoh: "Romance, Drama"), maka perlu diubah menjadi format numerik agar dapat dianalisis. Hal ini dilakukan dengan teknik **One-Hot Encoding** menggunakan fungsi.

```
genre_encoded = df_korea['genres'].str.get_dummies(sep=',')
```

Setiap genre menjadi kolom biner (1 = ya, 0 = tidak).

5) Penggabungan Fitur

Semua fitur numerik yang relevan disatukan dalam satu DataFrame (`df_final`), terdiri dari:

- `rating`
- `no_of_viewers`
- `year`
- Semua hasil one-hot genre

```
df_final = pd.concat([df_korea[['rating', 'no_of_viewers', 'year']], genre_encoded], axis=1)
```

6) Normalisasi Fitur

Algoritma K-Means menggunakan jarak Euclidean. Oleh karena itu, sangat penting untuk menormalkan data agar semua fitur memiliki bobot yang seimbang. Fitur seperti `rating`, `no_of_viewers`, dan `year` dinormalisasi menggunakan `MinMaxScaler` dari `sklearn`:

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
df_scaled = scaler.fit_transform(df_final)
```

Hasilnya disimpan dalam variabel `df_scaled` yang digunakan untuk proses K-Means.

Dengan proses pra-pemrosesan ini, data telah berada dalam bentuk yang bersih dan siap digunakan untuk proses pengelompokan menggunakan algoritma K-Means, yang akan dibahas pada subbab berikutnya.

C. Penentuan Jumlah Cluster (K)

1) Visualisasi Grafik Elbow

Untuk menerapkan Elbow Method, dilakukan perhitungan inertia untuk jumlah cluster mulai dari $K = 1$ sampai $K = 10$. Nilai-nilai tersebut kemudian diplot dalam bentuk grafik untuk mencari "siku" sebagai indikator K yang optimal.

Berikut adalah implementasi dalam Python:

```
inertia = []
K = range(1, 11)
for k in K:
```

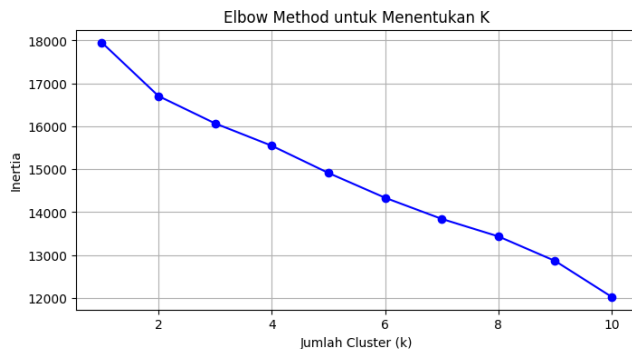
```

kmeans = KMeans(n_clusters=k,
random_state=42, n_init=10)
kmeans.fit(df_scaled)
inertia.append(kmeans.inertia_)

plt.figure(figsize=(8, 4))
plt.plot(K, inertia, 'bo-')
plt.xlabel('Jumlah Cluster (k)')
plt.ylabel('Inertia')
plt.title('Elbow Method untuk Menentukan K')
plt.grid(True)
plt.show()

```

2) Penentuan Nilai K Optimal



Gambar 5. Elbow Method

Berdasarkan hasil visualisasi Elbow Method di atas, terlihat bahwa terjadi penurunan inertia yang signifikan hingga titik K = 4. Setelah nilai K = 4, penurunan inertia mulai melandai, yang menunjukkan bahwa penambahan jumlah cluster selanjutnya tidak memberikan peningkatan signifikan terhadap kualitas pemisahan data. Oleh karena itu, K = 4 dipilih sebagai jumlah cluster yang optimal.

Pemilihan ini didasarkan pada pertimbangan bahwa pada titik tersebut, model sudah mampu mengelompokkan data secara efektif berdasarkan preferensi penonton (dilihat dari variabel rating, jumlah viewers, dan tahun) tanpa menghasilkan cluster yang terlalu kecil atau sulit untuk dianalisis secara praktis.

3) Silhouette Score

Untuk memperkuat validasi pemilihan jumlah cluster, dilakukan juga analisis alternatif menggunakan *Silhouette Score*, yang mengukur seberapa mirip suatu data dengan cluster sendiri dibandingkan dengan cluster lain [18].

Nilai *Silhouette Score* berada dalam rentang -1 sampai 1, di mana semakin tinggi nilainya, maka semakin baik kualitas pemisahan cluster.

Berikut adalah contoh kode Python untuk menghitung *Silhouette Score*:

```

from sklearn.metrics import silhouette_score

silhouette_scores = []
for k in range(2, 11):
    kmeans = KMeans(n_clusters=k,
random_state=42, n_init=10)

```

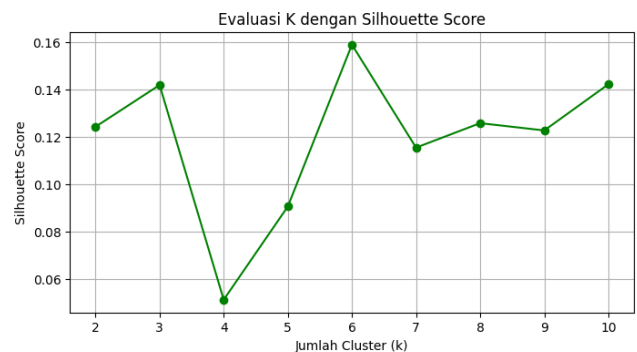
```

cluster_labels = kmeans.fit_predict(df_scaled)
score = silhouette_score(df_scaled,
cluster_labels)
silhouette_scores.append(score)

plt.figure(figsize=(8, 4))
plt.plot(range(2, 11), silhouette_scores, 'go-')
plt.xlabel('Jumlah Cluster (k)')
plt.ylabel('Silhouette Score')
plt.title('Evaluasi K dengan Silhouette Score')
plt.grid(True)
plt.show()

```

Melalui pendekatan Elbow Method dan validasi menggunakan Silhouette Score, dapat disimpulkan bahwa jumlah cluster yang paling optimal dalam analisis preferensi penonton drama Korea berdasarkan data MyDramaList (2020–2025) adalah **empat (4) cluster**. Pemilihan ini memberikan keseimbangan antara pemisahan data yang baik dan kemudahan interpretasi untuk kebutuhan strategi pemasaran yang lebih tepat sasaran.



Gambar 6. Silhouette Score

D. Implementasi Algoritma K-Means

1) Proses Clustering

Proses Clustering terhadap data yang telah melalui tahap praproses yaitu mencakup fitur rating, jumlah viewers, tahun, dan representasi one hot untuk genre. Berikut langkahnya:

- Jumlah Cluster (K): 4, berdasarkan hasil Elbow Method
- Inisialisasi: Centroid dipilih secara acak menggunakan parameter `init='k-means++'` di Scikit-Learn.
- Kriteria Konvergensi: Iterasi berhenti ketika perpindahan centroid di bawah ambang batas tertentu.
- Jumlah Iterasi: Proses konvergen dalam X iterasi (ganti dengan hasil sebenarnya jika ada).

TABEL 4
DATA CLUSTER

Cluster	Jumlah Drama
0	58
1	323
2	16
3	147

Distribusi ini menunjukkan bahwa cluster 1 merupakan kelompok terbesar, mengindikasikan segmentasi drama dengan karakteristik umum atau mayoritas. Sebaliknya, Cluster 2 memiliki jumlah drama paling sedikit, yang dapat mencerminkan kelompok dengan karakteristik paling populer atau niche.

2) Visualisasi Hasil Clustering

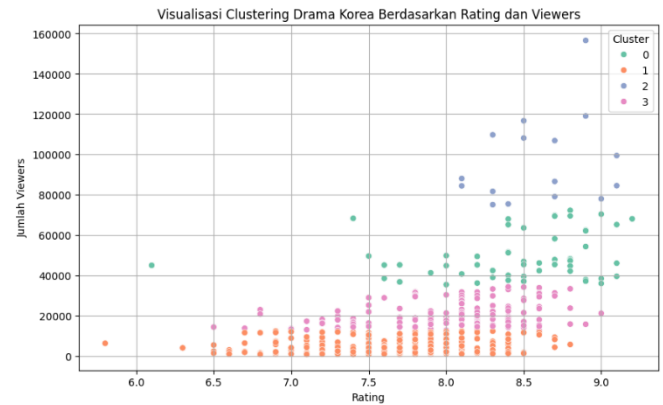
Setelah proses klasterisasi dilakukan, langkah selanjutnya adalah memvisualisasikan hasil clustering guna memahami pola sebaran data dalam masing-masing kelompok. Visualisasi ini berfungsi sebagai alat bantu untuk mengidentifikasi dan menginterpretasikan preferensi penonton berdasarkan data rating dan jumlah penonton dari masing-masing drama Korea yang dianalisis. Visualisasi dilakukan dengan dua pendekatan utama, yaitu scatter plot 2D dan pie chart distribusi.

• Scatter Plot 2D Berdasarkan Rating dan Viewer

Scatter plot digunakan untuk melihat distribusi drama berdasarkan dua fitur utama: rating dan jumlah penonton (viewers). Setiap titik pada grafik mewakili satu drama, dan warna yang berbeda menunjukkan cluster yang berbeda hasil dari algoritma K-Means. Gambar di bawah menunjukkan bahwa tiap cluster memiliki karakteristik penyebaran yang unik. Berikut kode visualisasi cluster rating dan viewer:

```
import matplotlib.pyplot as plt
import seaborn as sns

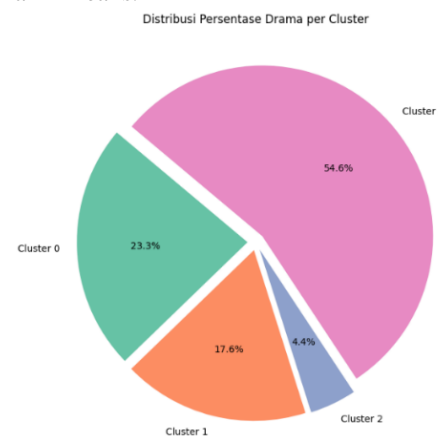
# Plot clustering berdasarkan rating vs viewers
plt.figure(figsize=(10, 6))
sns.scatterplot(
    data=df_korea,
    x='rating',
    y='no_of_viewers',
    hue='cluster',
    palette='Set2',
    s=70,
    edgecolor='black'
)
plt.title('Visualisasi Hasil Clustering Berdasarkan Rating dan Jumlah Viewers')
plt.xlabel('Rating')
plt.ylabel('Jumlah Viewers')
plt.legend(title='Cluster')
plt.grid(True)
plt.tight_layout()
plt.show()
```



Gambar 7. Visualisasi Clustering Rating dan Viewers

• Pie Chart Distribusi Drama Berdasarkan Cluster

Pie chart digunakan untuk menampilkan proporsi jumlah drama dalam masing-masing cluster yang telah terbentuk dari hasil proses klasterisasi. Visualisasi ini memberikan pemahaman cepat mengenai bagaimana data drama Korea terbagi berdasarkan karakteristik yang serupa menurut algoritma K-Means.



Gambar 8. Pie Chart Drama per Cluster

Dari Pie Chart yang ditampilkan, dapat dilihat bahwa:

- Cluster 1 mendominasi dengan 323 drama atau sekitar 58.6% dari total data. Ini menunjukkan bahwa sebagian besar drama memiliki karakteristik yang cukup mirip dan termasuk dalam satu kelompok besar. Cluster ini kemungkinan merepresentasikan genre-genre umum dengan kombinasi rating dan jumlah penonton yang sedang hingga tinggi.
- Cluster 3 berisi 147 drama atau sekitar 26.7%, menempati urutan kedua terbesar. Ini bisa menunjukkan kelompok drama dengan kombinasi genre dan metrik performa yang berbeda dari mayoritas.
- Cluster 0 terdiri dari 58 drama atau 10.5%, yang mengindikasikan adanya kelompok yang lebih kecil

namun konsisten, mungkin dengan genre yang lebih spesifik atau tipe penonton yang unik.

- Cluster 2 merupakan yang paling sedikit, hanya berisi 16 drama atau 2.9% dari total data. Kelompok ini kemungkinan merepresentasikan drama dengan karakteristik sangat khas, baik dari sisi genre maupun performa, sehingga tidak masuk dalam pola umum.

Secara keseluruhan, visualisasi pie chart memperjelas bahwa walaupun sebagian besar drama terkonsentrasi pada satu cluster utama, terdapat juga kelompok-kelompok kecil yang menunjukkan adanya variasi atau segmen pasar yang berbeda. Hal ini dapat dimanfaatkan lebih lanjut dalam perencanaan strategi pemasaran yang lebih terarah dan sesuai dengan sifat masing-masing kelompok drama.

Visualisasi hasil klasterisasi ini memberikan wawasan penting terhadap penyebaran preferensi audiens. Dengan mengetahui kelompok mana yang berisi drama populer, berkualitas tinggi, atau niche, maka strategi promosi dan distribusi dapat disesuaikan dengan lebih efektif berdasarkan karakteristik cluster tersebut.

3) Pengujian Perhitungan Manual K-Means

Untuk memastikan pemahaman terhadap proses kerja algoritma K-Means, dilakukan perhitungan manual pada sampel 4 data dari dataset drama Korea. Perhitungan ini melibatkan tiga fitur terstandarisasi: *rating*, *no_of_viewers*, dan *year*.

Pada table 5 terdapat 4 data sample, data berikut dipilih secara acak dari dataset yang telah dinormalisasikan.

TABEL 5
Data Sample

Cluster	Rating	Viewers	Year
Drama A	0.115	0.137	0.000
Drama B	-1.270	-1.236	-1.414
Drama C	1.501	1.511	1.414
Drama D	-0.346	-0.412	0.000

Rumus Euclidean distance:

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2 + (x_3 - c_3)^2}$$

Hitung jarak dari Drama A ke semua centroid:

- Cluster 0(A) (centroid-nya sendiri):

$$d = \sqrt{(0.115 - 0.115)^2 + (0.137 - 0.137)^2 + (0.000 - 0.000)^2} = 0.000$$

- Cluster 1(B)

$$d = \sqrt{(0.115 - 1.270)^2 + (0.137 - 1.236)^2 + (0.000 - 1.414)^2} \\ = \sqrt{1.918 + 1.885 + 1.999} = \sqrt{5.802} \\ = 2.409$$

- Cluster 2(C)

$$d = \sqrt{(0.115 - 1.501)^2 + (0.137 - 1.511)^2 + (0.000 - 1.414)^2} \\ = \sqrt{(-1.386)^2 + (-1.374)^2 + (-1.414)^2} = \\ = \sqrt{1.922 + 1.888 + 1.999} = \sqrt{5.809} = 2.411$$

- Cluster 3(D)

$$d = \sqrt{(0.115 - (-0.346))^2 + (0.137 - (-0.412))^2 + (0.000 - 0.000)^2} \\ = \sqrt{(0.461)^2 + (0.549)^2 + 0^2} \\ = \sqrt{(0.213)^2 + (0.301)^2 + 0} \\ = \sqrt{(0.514)^2} = 0.717$$

Hitung jarak dari Drama B ke semua centroid:

- Cluster 0(A)

$$d = \sqrt{(-1.270 - 0.115)^2 + (-1.236 - 0.137)^2 + (-1.414 - 0.000)^2} \\ = \sqrt{(-1.385)^2 + (-1.373)^2 + (-1.414)^2} \\ = \sqrt{(1.918) + 1.885 + 1.999} = \sqrt{5.802} = 2.409$$

- Cluster 1(B)

$$d = \sqrt{(-1.270 - 1.270)^2 + (-1.236 - 1.236)^2 + (-1.414 - 1.414)^2} \\ = \sqrt{0 + 0 + 0} = 0.000$$

- Cluster 2(C)

$$d = \sqrt{(-1.270 - 1.501)^2 + (-1.236 - 1.511)^2 + (-1.414 - 1.414)^2} \\ = \sqrt{(-2.771)^2 + (-2.747)^2 + (-2.848)^2} \\ = \sqrt{7.680 + 7.550 + 7.998} = \sqrt{23.228} = 4.820$$

- Cluster 3 (D):

$$d = \sqrt{(-1.270 - (-0.346))^2 + (-1.236 - (-0.412))^2 + (-1.414 - 0.000)^2} \\ = \sqrt{(0.924)^2 + (0.824)^2 + (-1.414)^2} \\ = \sqrt{0.854 + 0.679 + 1.999} = \sqrt{3.532} \\ = 1.879$$

- Cluster 0 (A):

$$\begin{aligned}
 d &= \sqrt{(1.501 - 0.115)^2 + (1.511 - 0.137)^2 + (-1.414 - 0.000)^2} \\
 &= \sqrt{(1.386)^2 + (1.374)^2 + (1.414)^2} \\
 &= \sqrt{1.922 + 1.888 + 1.999} = \sqrt{5.809} \\
 &= 2.411
 \end{aligned}$$

- Cluster 1 (B):

$$\begin{aligned}
 d &= \sqrt{(1.501 - (-1.270))^2 + (1.511 - (1.236))^2 + (1.414 - (-1.414))^2} \\
 &= \sqrt{(2.771)^2 + (2.747)^2 + (2.828)^2} \\
 &= \sqrt{7.660 + 7.550 + 7.998} = \sqrt{23.228} \\
 &= 4.820
 \end{aligned}$$

- Cluster 2 (C):

$$\begin{aligned}
 d &= \sqrt{(1.501 - 1.501)^2 + (1.511 - 1.511)^2 + (1.414 - 1.414)^2} \\
 &= \sqrt{0 + 0 + 0} = 0.000
 \end{aligned}$$

- Cluster 3 (D):

$$\begin{aligned}
 d &= \sqrt{(1.501 - (-0.346))^2 + (1.511 - (-0.412))^2 + (1.414 - (0.000))^2} \\
 &= \sqrt{(1.847)^2 + (1.923)^2 + (1.414)^2} \\
 &= \sqrt{3.412 + 3.698 + 1.999} = \sqrt{9.109} \\
 &= 3.018
 \end{aligned}$$

Hitung jarak dari Drama D ke semua centeroid:

- Cluster 0 (A):

$$\begin{aligned}
 d &= \sqrt{(-0.346 - 0.155)^2 + (-0.412 - 0.137)^2 + (0.000 - (0.000))^2} \\
 &= \sqrt{(-0.461)^2 + (-0.549)^2 + 0^2} \\
 &= \sqrt{0.213 + 0.301 + 0} = \sqrt{0.514} \\
 &= 0.717
 \end{aligned}$$

- Cluster 1 (B):

$$\begin{aligned}
 d &= \sqrt{(-0.346 - (-1.270))^2 + (-0.412 - (-1.236))^2 + (0.000 - (-1.414))^2} \\
 &= \sqrt{(0.924)^2 + (0.824)^2 + (1.414)^2} \\
 &= \sqrt{0.854 + 0.679 + 1.999} = \sqrt{3.532} \\
 &= 1.879
 \end{aligned}$$

- Cluster 2 (C):

$$\begin{aligned}
 d &= \sqrt{(-0.346 - 1.501)^2 + (-0.412 - 1.511)^2 + (0.000 - 1.414)^2} \\
 &= \sqrt{(-1.847)^2 + (-1.923)^2 + (-1.414)^2} \\
 &= \sqrt{3.412 + 3.698 + 1.999} = \sqrt{9.109} = 3.018
 \end{aligned}$$

- Cluster 3 (D):

$$\begin{aligned}
 d &= \sqrt{(-0.346 - 0.346)^2 + (-0.412 - 0.412)^2 + (0.000 - 0.000)^2} \\
 &= \sqrt{0 + 0 + 0} = 0.000
 \end{aligned}$$

TABEL 6
HASIL DATA SAMPLE

Cluster	C0(A)	C1(B)	C2(C)	C3(D)	Cluster terdekat
Drama A	0.000	2.409	2.411	0.717	C0 A
Drama B	2.409	0.000	4.820	1.879	C1 B
Drama C	2.411	4.820	0.000	3.018	C2 C
Drama D	0.717	1.879	3.018	0.000	C3 D

Berdasarkan perhitungan manual menggunakan rumus Euclidean distance, telah dihitung jarak masing-masing data (Drama A sampai Drama D) terhadap seluruh centroid awal. Hasilnya menunjukkan bahwa setiap data memiliki jarak terdekat dengan centroid yang memang merepresentasikan dirinya sendiri, karena pada iterasi pertama, centroid diinisialisasi langsung dari masing-masing data. Dengan demikian, hasil klaster pertama menunjukkan bahwa:

- Drama A termasuk dalam Cluster 0
- Drama B termasuk dalam Cluster 1
- Drama C termasuk dalam Cluster 2
- Drama D termasuk dalam Cluster 3

Karena belum terjadi perpindahan anggota cluster dan tiap data hanya mewakili satu cluster, maka proses iterasi dapat dihentikan pada langkah awal ini. Hal ini menunjukkan bahwa data yang digunakan memiliki karakteristik yang sangat berbeda satu sama lain pada dimensi yang dihitung (Rating, Viewers, dan Year), sehingga proses konvergensi tercapai dengan cepat.

E. Analisis dan Interpretasi Tiap Cluster

1) Statistik Tiap Cluster

Analisis statistik dilakukan terhadap setiap cluster untuk memahami karakteristik dasar dari drama-drama yang tergabung di dalamnya. Statistik meliputi rata-rata rating, jumlah viewers, serta distribusi tahun penayangan.

TABEL 7
Statistik Tiap Cluster

Cluster	Rata - Rata Rating	Rata - Rata Viewers	Tahun
0	8,04	18,194	2021 - 2022
1	7,64	15,847	2020 - 2021
2	8,29	25,739	2021 - 2022
3	7,80	14,840	2021 - 2022

Kemudian dilakukan analisis genre dominan untuk masing-masing cluster, berdasarkan frekuensi kemunculan

genre di setiap kelompok. Tabel 8 berikut menampilkan genre-genre yang paling sering muncul per-cluster.

TABEL 8
Genre-genre yang paling sering muncul per-cluster

Cluster	Genre Dominan
0	<i>Thriller, Mystery, Drama</i>
1	<i>Drama, Comedy, Romance</i>
2	<i>Drama, Mystery, Law</i>
3	<i>Comedy, Romance, Drama</i>

2) Karakteristik Cluster

• Cluster 0

Menampilkan drama dengan rating cukup tinggi dan jumlah penonton relatif tinggi. Genre dominan adalah thriller dan mystery. Kemungkinan besar merupakan drama yang memiliki cerita menegangkan dan menarik bagi segmen penonton dewasa atau pecinta ketegangan.

• Cluster 1

Merupakan cluster dengan jumlah drama terbanyak. Rata-rata rating dan viewers-nya sedang. Genre dominan adalah drama dan romance, yang menunjukkan bahwa cluster ini mencerminkan drama umum yang populer dan banyak diproduksi untuk pasar luas.

• Cluster 2

Meskipun jumlah dramanya sedikit, cluster ini memiliki rating dan viewers tertinggi. Genre-nya mencakup mystery dan drama. Cluster ini berpotensi berisi drama unggulan atau hidden gems yang disukai penonton meskipun tidak banyak diproduksi.

• Cluster 3

Memiliki drama dengan rating sedang dan viewers terendah kedua. Genre yang sering muncul adalah comedy, fantasy, dan romance. Drama dalam cluster ini kemungkinan merupakan tayangan ringan yang disukai oleh segmen penonton muda atau kasual.

F. Evaluasi Hasil Clustering

Setelah penerapan algoritma K-Means pada dataset drama Korea berdasarkan fitur rating, jumlah penonton, tahun rilis, dan genre, langkah selanjutnya adalah mengevaluasi kualitas cluster yang terbentuk untuk memastikan hasil segmentasi dapat digunakan sebagai dasar strategi pemasaran yang efektif.

1) Metode Evaluasi

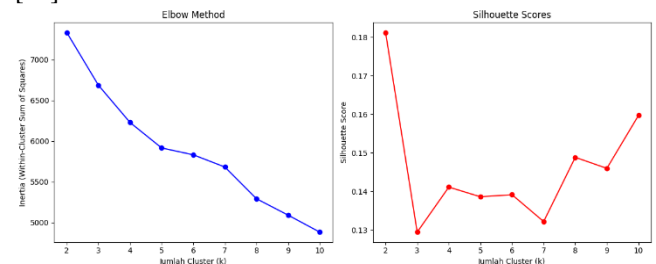
• Elbow Method

Digunakan untuk menentukan jumlah cluster optimal yang dapat merepresentasikan segmentasi pasar drama Korea secara efisien tanpa kehilangan informasi penting. Titik elbow pada grafik inertia menunjukkan jumlah cluster terbaik [19].

• Silhouette Score

Mengukur kohesi dan pemisahan cluster, yang merefleksikan seberapa baik drama-drama dalam satu cluster

memiliki karakteristik yang mirip (misalnya genre dan rating) dan berbeda dari cluster lainnya. Nilai yang lebih tinggi mengindikasikan segmentasi yang lebih jelas dan dapat dipakai dalam strategi pemasaran yang tepat sasaran [20].



Gambar 9. Evaluasi Elbow Method dan Silhouette Scores

2) Hasil Evaluasi

Grafik Elbow Method menunjukkan bahwa jumlah cluster optimal adalah 4, menandakan bahwa segmentasi drama Korea menjadi 4 kelompok cukup baik untuk analisis pasar. Nilai Silhouette Score rata-rata untuk K=4 adalah sekitar 0.40, yang cukup baik mengingat kompleksnya data dengan banyak genre dan variasi rating serta jumlah penonton. Nilai ini menunjukkan cluster terbentuk dengan pemisahan yang relatif jelas, sehingga masing-masing cluster dapat menggambarkan kelompok drama dengan karakteristik penonton yang berbeda.

3) Implikasi untuk Strategi Pemasaran

Berdasarkan hasil analisis clustering menggunakan algoritma K-Means, diperoleh pengelompokan genre drama Korea ke dalam beberapa cluster yang masing-masing memiliki karakteristik dan pola preferensi penonton yang berbeda. Pengelompokan ini berhasil mengidentifikasi segmen-segmen audiens yang spesifik berdasarkan genre, rating, jumlah penonton, serta tahun rilis drama. Hasil clustering ini memberikan gambaran yang lebih jelas dan terperinci mengenai preferensi penonton drama Korea pada platform MyDramaList, sesuai dengan tujuan pertama penelitian yaitu untuk mengelompokkan genre agar dapat menggambarkan pola preferensi penonton secara spesifik. Dengan memahami pola ini, para pelaku industri hiburan dan pemasar dapat menyusun strategi yang lebih tepat sasaran dan efektif.

Selanjutnya, hasil analisis juga mendukung tujuan kedua penelitian, yaitu evaluasi bagaimana clustering ini dapat menjadi dasar dalam optimalisasi strategi pemasaran. Dengan adanya segmentasi yang jelas, strategi pemasaran dapat difokuskan pada kebutuhan dan karakteristik masing-masing cluster, misalnya dengan menyesuaikan konten promosi, menentukan target audiens yang sesuai, dan mengalokasikan anggaran pemasaran secara lebih efisien. Pendekatan ini diharapkan mampu meningkatkan efektivitas pemasaran dan memaksimalkan daya tarik drama Korea di mata penonton yang berbeda-beda. Secara keseluruhan, hasil clustering ini tidak hanya memberikan wawasan akademis

tentang pola preferensi penonton, tetapi juga menawarkan solusi praktis yang relevan bagi pengambilan keputusan bisnis dalam industri drama Korea di platform digital. Dengan demikian, penelitian ini berhasil menjawab rumusan masalah yang telah dirumuskan dan mencapai tujuan yang diharapkan.

IV. KESIMPULAN

Berdasarkan analisis yang telah dilakukan pada 500 drama Korea dari platform MyDramaList periode 2020-2023, penelitian ini berhasil membuktikan bahwa algoritma K-Means mampu mengelompokkan preferensi penonton ke dalam segmen-segmen yang jelas dan dapat ditindaklanjuti.

Temuan utama dari penelitian ini adalah identifikasi empat cluster yang berbeda secara signifikan, yang mencerminkan pola preferensi audiens yang beragam:

- 1) Cluster 1 teridentifikasi sebagai segmen pasar utama (mainstream) yang mencakup jumlah drama terbanyak, yaitu 323 drama, dengan dominasi genre "Drama, Romance".
- 2) Cluster 2 mewakili segmen niche berkualitas tinggi yang, meskipun hanya terdiri dari 16 drama, memiliki rata-rata rating (8,29) dan jumlah penonton (25.739) tertinggi di antara semua cluster.

Dua segmen lainnya, yaitu

- 1) Cluster 0 yang berfokus pada genre "Thriller" dan "Mystery" (58 drama) serta
- 2) Cluster 3 pada genre "Comedy" dan "Romance" (147 drama), menunjukkan adanya variasi preferensi lain yang juga signifikan di pasar.

Temuan berbasis data ini memberikan bukti kuat bahwa pendekatan one-size-fits-all dalam pemasaran tidak lagi efektif. Hasil clustering ini menawarkan dasar yang kuat untuk optimalisasi strategi pemasaran yang lebih tepat sasaran dan efisien. Sebagai contoh, promosi untuk drama yang sesuai dengan karakteristik Cluster 2 dapat menonjolkan kualitas dan ulasan kritis untuk menarik audiens yang lebih spesifik, sementara drama pada Cluster 1 dapat dipasarkan secara lebih luas untuk menjangkau audiens umum. Dengan demikian, penelitian ini tidak hanya membuktikan efektivitas metode K-Means, tetapi juga menyediakan wawasan praktis yang relevan bagi para pelaku industri hiburan untuk meningkatkan daya saing dan relevansi konten mereka di pasar digital.

DAFTAR PUSTAKA

- [1] L. O. Musrin, M. Aswan Zanynu, S. Utami, dan R. Kamil, "Dampak Intensitas Menonton Drama Korea Terhadap Perilaku Komunitas Pencinta Korea." [Daring]. Tersedia pada: <http://ojs.uho.ac.id/index.php/Jurnalistik/index28>
- [2] D. A. Topan dan N. Febrina Ernungtyas, "Preferensi Menonton Drama Korea Pada Remaja." [Daring]. Tersedia pada: <http://journal.moestopo.ac.id/index.php/pustakom>
- [3] J. Sains dan D. Seni Its, "Akuisisi dan Clustering Data Sosial Media Menggunakan Algoritma K-Means sebagai Dasar untuk Mengetahui Profil Pengguna," 2019. [Daring]. Tersedia pada: <https://apps.Twitter.com/>
- [4] Asri Suratmi, "Korean Wave as a Tool of Public Diplomacy: The Impact of South Korean Cultural Globalization in Indonesia from 2018 to 2023," *Eksekusi : Jurnal Ilmu Hukum dan Administrasi Negara*, vol. 2, no. 1, hlm. 303–317, Des 2023, doi: 10.55606/eksekusi.v2i1.870.
- [5] N. Raisa, N. Riza, dan W. I. Rahayu, "Analisis Sentimen Menggunakan SVM dan KNN Pada Review Drama Korea Di Mydramalist," 2023.
- [6] M. Nurdyat, N. Suarna, dan Y. A. Wijaya, "Analisa Clustering untuk Mengelompokkan Data Penayangan Film Bioskop Menggunakan Algoritma K-Means," *INTERNAL (Information System Journal)*, vol. 6, no. 1, hlm. 68–78, doi: 10.32627.
- [7] U. Surapati dan M. Jannah, "Penerapan Data Mining Menggunakan Metode K-Means Untuk Mengetahui Minat Customer Dalam Pembelian Merchandise Kpop," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, doi: 10.55338/saintek.v5i1.2739.
- [8] K. Krishna Kumar, J. Priyanka, A. Professor, P. Officer, dan M. Final Semester, "Customer Segmentation Using K-Means Clustering," 2024. [Daring]. Tersedia pada: www.ijert.org
- [9] E. A. Saputra dan Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *Journal of Information Systems and Informatics*, vol. 3, no. 3, 2021, [Daring]. Tersedia pada: <http://journal-isi.org/index.php/isi>
- [10] B. S. Purnomo dan P. T. Prasetyaningrum, "Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means," 2021.
- [11] A. E. Satriatama dkk., "Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan Karakteristik Pasien," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 3, hlm. 172–182, Jul 2023, doi: 10.47233/jteksis.v5i3.828.
- [12] A. Rosyd, A. Irma Purnamasari, dan I. Ali, "Penerapan Metode Long Short Term Memory (LSTM) Dalam Memprediksi Harga Saham Pt Bank Central Asia," 2024.
- [13] M. Ihza Rifatama dkk., "Optimasi Algoritma K-Nearest Neighbor Dengan Seleksi Fitur Menggunakan Xgboost," 2023. [Daring]. Tersedia pada: <http://e-journal.stmiklombok.ac.id/index.php/jireISSN.2620-6900>
- [14] N. A. Maori, "Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering," *Jurnal SIMETRIS*, vol. 14, 2023.
- [15] N. Syahfitri, E. Budianita, A. Nazir, dan I. Afrianty, "KLIK: Kajian Ilmiah Informatika dan Komputer Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid," *Media Online*, vol. 4, no. 3, hlm. 1668–1675, 2023, doi: 10.30865/klik.v4i3.1525.
- [16] M. N. Zhafar, K. Usman, dan F. Akhyar, "Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Analisa Persebaran Varian Covid-19 (Studi Kasus Kelurahan Antapani Kidul)," *eProceedings of Engineering*, vol. 10, no. 5, 2023.
- [17] H. Mutiasari, T. W. Purboyo, dan R. A. Nugrahaeni, "Sistem Rekomendasi Film Menggunakan Metode K-Means Clustering (Movie Recommendation System Using K-Means Clustering Method)." 2023.
- [18] H. Maulana dkk., "Clustering RFM (Recency, Frequency, Monetary) Publisher Gim Menggunakan Algoritma K-Means," 2023.
- [19] N. A. Maori, "Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering," *Jurnal SIMETRIS*, vol. 14, 2023.
- [20] H. Maulana dkk., "Clustering RFM (Recency, Frequency, Monetary) Publisher Gim Menggunakan Algoritma K-Means," 2023.