PENERAPAN ALGORITMA K-MEANS UNTUK KLASTERISASI FILM PADA PLATFORM AMAZON PRIME VIDEO

Naufal Fauzi Rahman, Bayu Priyatna, Fitria Nurapriani, Tukino Paryono
Sistem Informasi, Universitas Buana Perjuangan Karawang
Jl. HS.Ronggo Waluyo, Puseurjaya, Telukjambe Timur,
Karawang, Jawa Barat 41361, Indonesia
si23.naufalrahman@mhs.ubpkarawang.ac.id

ABSTRAK

Pesatnya perkembangan platform *streaming* seperti Amazon Prime Video telah menghasilkan volume data film yang sangat besar, sehingga membuka peluang untuk analisis *data mining*. Masalah yang muncul adalah sulitnya memahami struktur pasar dari katalog film yang sangat besar tanpa segmentasi yang jelas. Oleh karena itu, studi ini ditujukan untuk mengimplementasikan algoritma K-Means untuk klasterisasi film yang didasari empat atribut utama yaitu, rating, popularitas, tahun rilis, dan harga. Metode penelitian mengadopsi kerangka *Knowledge Discovery in Databases* (KDD), di mana jumlah *cluster* yang optimal ditentukan menggunakan *Elbow Method* dan kualitasnya dievaluasi menggunakan *Silhouette Score*. Hasil analisis berhasil mengelompokkan 2.108 data film ke dalam 6 *cluster* yang valid dan berbeda secara signifikan. Segmen yang diidentifikasi meliputi "Blockbuster Sangat Populer & Premium", "Film Modern Berkualitas Tinggi", dan "Film Klasik & Lama". Temuan ini memberikan pemetaan segmen pasar yang jelas dan dapat digunakan sebagai dasar untuk personalisasi rekomendasi dan strategi pemasaran konten yang lebih efektif.

Kata kunci : *Klasterisasi, K-Means, Data Mining, Segmentasi Film, Amazon Prime Video, Sistem Rekomendasi*

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong pertumbuhan eksponensial volume data di berbagai sektor, fenomena yang dikenal sebagai *Big Data* [1]. Salah satu sektor yang mengalami dampak signifikan adalah industri hiburan digital, khususnya layanan streaming *video-on-demand* (VOD) seperti Amazon Prime Video, Netflix, dan Disney+. Platform-platform ini tidak hanya berfungsi sebagai media penyiaran tetapi juga sebagai generator data besar yang mencatat preferensi, perilaku, dan interaksi jutaan pengguna setiap hari [2]. Data yang melimpah ini memiliki potensi besar untuk dieksplorasi guna mendapatkan wawasan berharga yang dapat meningkatkan kualitas layanan dan strategi bisnis.

Untuk mengubah data mentah menjadi informasi yang bermakna, diperlukan pendekatan analitis yang sistematis. Di sinilah *data mining* memainkan peran krusial. *Data Mining merupakan* proses menemukan pola atau pengetahuan menarik dari dataset besar dengan mengikuti tahap-tahap seperti *Knowledge Discovery in Databases* (KDD) atau *Cross-Industry Standard Process for Data Mining* (CRISP-DM) [3]. Teknik ini memungkinkan organisasi untuk mengidentifikasi tren tersembunyi, memahami perilaku pelanggan, dan membuat keputusan berbasis data yang lebih akurat. Salah satu teknik dasar dan kuat dalam *data mining* adalah metode *clustering*.

Clustering adalah metode analisis data untuk mengelompokkan sekumpulan objek ke dalam beberapa kelompok atau *cluster* [4]. Objek dalam satu *cluster* memiliki tingkat kesamaan yang tinggi satu sama lain, sementara objek dalam *cluster* yang berbeda memiliki tingkat kesamaan yang rendah [5]. Dalam konteks layanan *streaming*, *clustering* dapat digunakan untuk mengelompokkan film, serial TV, atau bahkan pengguna berdasarkan atribut tertentu. Hasil pengelompokan ini dapat dimanfaatkan untuk berbagai tujuan, seperti personalisasi sistem rekomendasi [6] dan segmentasi pasar untuk strategi pemasaran yang lebih tertarget [7].

Di antara berbagai algoritma *clustering*, algoritma K-Means adalah salah satu metode yang paling populer dan banyak diterapkan karena kesederhanaan konsepnya, efisiensi komputasi, dan kemudahan dalam menafsirkan hasil [8]. K-Means bekerja dengan membagi data menjadi K *cluster* yang telah ditentukan sebelumnya, di mana setiap titik data menjadi anggota *cluster* dengan pusat *cluster* (*centroid*) terdekat [9]. Algoritma ini telah terbukti efektif dalam berbagai studi kasus, mulai dari segmentasi data pelanggan hingga pengelompokan menu makanan berdasarkan kebutuhan kalori [10].

Meskipun Amazon Prime Video memiliki katalog film yang luas, sedikit analisis yang secara khusus mengelompokkan katalog tersebut untuk memahami struktur pasarnya. Namun, memahami segmen film yang ada seperti film *blockbuster* populer, film klasik berkualitas tinggi, atau film *niche* dapat memberikan wawasan penting bagi platform. Oleh karena itu, tujuan dari penelitian ini adalah untuk menggunakan algoritma K-Means dalam mengkategorikan kumpulan data film yang dapat diakses di platform Amazon Prime Video. Pengelompokan akan didasarkan pada empat atribut kunci yaitu, peringkat film, popularitas (jumlah peringkat), tahun rilis, dan harga. Diharapkan hasil penelitian ini dapat mengidentifikasi dan memprofilkan segmen film yang secara alami muncul dari data [11], yang mampu memberikan kontrbusi

untuk merancang strategi konten dan pemasaran yang lebih efisien serta berdampak.

2. TINJAUAN PUSTAKA

Bagian ini membahas landasan teoritis penelitian, meliputi prinsip *data mining*, klasterisasi, algoritma K-Means, dan analisis berbagai penelitian terkait yang telah dilakukan sebelumnya.

2.1. Data Mining

Data mining merupakan suatu area yang bertujuan untuk memperoleh informasi krusial serta pola-pola yang tidak terlihat dari sejumlah besar data [12]. Proses ini tidak hanya tentang pengumpulan data, tetapi serangkaian langkah terstruktur untuk mengonversi informasi dasar menjadi wawasan yang bisa diterapkan. Salah satu metodologi yang paling umum diterapkan dalam proyek-proyek *data mining* adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Metode ini mencakup beberapa tahap penting, antara lain pemahaman terhadap aspek bisnis, eksplorasi data, pengolahan data, pembuatan model, penilaian, dan akhirnya penerapan [3]. Melalui implementasi *data mining*, organisasi dapat memperoleh keunggulan kompetitif dengan memahami tren pasar dan perilaku konsumen secara lebih mendalam.

2.2. Klasterisasi (Clustering)

Klasterisasi merupakan salah satu tugas utama dalam *data mining* yang termasuk dalam kategori pembelajaran tanpa pengawasan (*unsupervised learning*) [8]. Inti dari klasterisasi adalah mengatur data ke dalam beragam kelompok atau *cluster* dengan cara yang membuat data di dalam satu *cluster* menunjukkan tingkat kesamaan yang sangat tinggi. Sebaliknya, data yang berada di dalam *cluster* yang berbeda seharusnya menunjukkan tingkat kesamaan yang lebih rendah [5]. Teknik ini sangat berguna ketika tidak ada label atau kategori yang telah ditentukan sebelumnya dalam data. Dalam praktiknya, *clustering* sering digunakan untuk segmentasi pelanggan, deteksi anomali, analisis gambar, dan segmentasi pasar [7].

2.3. Algoritma K-Means

Metode klasterisasi K-Means adalah salah satu algoritma non-hierarkis yang paling terkenal dan sering diterapkan dalam berbagai bidang. Algoritma ini dirancang untuk memisahkan n observasi data menjadi k *cluster*, di mana setiap observasi menjadi bagian dari *cluster* yang memiliki rata-rata atau pusat yang paling dekat [9]. Algoritma K-Means umumnya melibatkan langkah-langkah berikut:

a. Tentukan berapa jumlah *cluster* (k) yang ingin dibentuk.
b. Lakukan inisialisasi dengan memilih k titik *centroid* secara acak sebagai titik pusat awal untuk masing-masing *cluster*.
c. Hitunglah jarak dari setiap titik data ke masing-masing *centroid*. Kemudian, tetapkan setiap titik data ke *cluster* yang memiliki *centroid* terdekat.
d. Perbarui posisi *centroid* setiap *cluster* dengan menghitung rata-rata dari semua titik data yang telah ditugaskan ke *cluster* tersebut.
e. Ulangi langkah 3 dan 4 hingga posisi *centroid cluster* tidak mengalami perubahan yang signifikan lagi, atau hingga tercapai jumlah iterasi maksimum yang telah ditentukan.

Keuntungan utama K-Means adalah relatif mudah diimplementasikan dan cepat dalam hal perhitungan. Namun, algoritma ini juga memiliki kelemahan, yaitu hasilnya dapat sensitif terhadap pemilihan titik pusat awal (*centroid*) dan kebutuhan untuk menentukan jumlah *cluster* (k) secara manual. Untuk menyelesaikan isu terkait penentuan nilai k, teknik seperti *Elbow Method* sering kali diterapkan guna mengidentifikasi jumlah *cluster* yang paling efisien [10].

2.4. Penelitian Terkait

Studi sebelumnya telah menerapkan algoritma K-Means untuk analisis dan segmentasi data film di berbagai platform. Penelitian Suarna menerapkan K-Means untuk mengelompokkan data film di platform Netflix. Dalam studi ini, metode KDD (*Knowledge Discovery in Database*) digunakan sebagai kerangka kerja, dan hasil pengelompokan berhasil mengidentifikasi pola film berdasarkan atribut spesifik yang tersedia di Netflix. Demikian pula, Fitrianti juga melakukan klasterisasi pada film-film populer di Netflix menggunakan kerangka kerja CRISP-DM, di mana tiga atribut utama yaitu, durasi, rating, dan jumlah suara digunakan untuk membentuk tiga klaster film yang berbeda.

Budihartanti juga meneliti data film Netflix dengan tujuan memberikan rekomendasi. Studi ini menggunakan K-Means untuk mengelompokkan film dan berhasil menunjukkan bahwa metode ini efektif untuk segmentasi data sebagai dasar sistem rekomendasi. Di platform berbeda, Ashari menerapkan K-Means pada data film dari IMDb, menggunakan *Davies Bouldin Index* berfungsi sebagai alat evaluasi guna menjamin mutu dari kluster yang terbentuk.

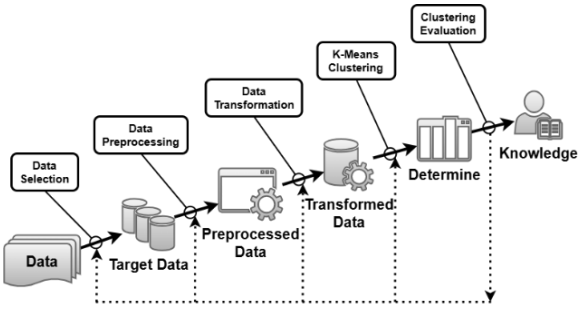
Dalam konteks lokal, Vania menggunakan K-Means untuk merekomendasikan film Indonesia, menunjukkan kesuksesan algoritma dalam mengelompokkan film ke dalam kategori "direkomendasikan" dan "kurang direkomendasikan". Di luar domain film, Sasmoyo menunjukkan fleksibilitas K-Means dengan menggunakannya untuk memetakan area yang rentan terhadap kematian anak di Jawa Barat, menunjukkan kemampuan algoritma dalam melakukan segmentasi geografis.

Dari tinjauan di atas, jelas bahwa penerapan K-Means untuk pengelompokan data film, khususnya di platform Netflix, telah dilakukan secara luas. Namun, penelitian yang secara khusus berfokus pada katalog

Amazon Prime Video menggunakan kombinasi empat atribut (peringkat, popularitas, tahun rilis, dan harga) untuk segmentasi pasar masih memiliki ruang untuk eksplorasi lebih lanjut. Sehubungan dengan hal tersebut, penelitian ini memiliki tujuan untuk menjembatani kekurangan yang ada dengan menerapkan analisis serupa pada dataset yang berbeda untuk menghasilkan wawasan baru.

3. METODE PENELITIAN

Metode dalam penelitian ini memanfaatkan kerangka kerja *Knowledge Discovery in Databases* (KDD), yang merupakan sebuah proses sistematis untuk menguak pengetahuan dari kumpulan data. Proses KDD terdiri dari beberapa tahapan penting, yang mencakup pemilihan data, pembersihan dan transformasi data, penerapan teknik *data mining*, dan juga evaluasi serta penafsiran hasil yang diperoleh [2]. Alur penelitian digambarkan melalui Gambar 1. dibawah ini.



Gambar 1. Tahapan Proses KDD

3.1. Seleksi Data

Tahap awal studi ini adalah seleksi data. Data yang digunakan adalah dataset publik berjudul "Dataset Film Amazon dari Tahun 1931 - 2023". Dataset ini berisi 2.108 data film (baris) dengan 10 atribut (kolom). Dari semua atribut yang tersedia, dilakukan pemilihan untuk memilih fitur yang relevan dengan tujuan pengelompokan. Fitur yang dipilih adalah fitur numerik yang dapat mewakili karakteristik utama suatu film, yaitu:

a. Movie_Rating: Rating film yang diberikan oleh pengguna.
b. No_of_Ratings: Jumlah total pengguna yang memberikan rating, yang mewakili tingkat popularitas.
c. ReleaseYear: Tahun film dirilis.
d. Price: Harga sewa atau beli film dalam Dolar AS.

3.2. Pra-Pemrosesan dan Transformasi Data

Setelah diseleksi, data mentah tersebut melalui tahap *pre-processing* dan transformasi untuk memastikan kualitas dan kesiapannya sebelum diolah oleh algoritma K-Means.

a. Penanganan Data Hilang: Dilakukan pemeriksaan terhadap nilai-nilai kosong (*null* atau NaN) pada fitur yang dipilih. Ditemukan bahwa atribut ReleaseYear dan Price terdapat data yang hilang. Untuk mengatasi hal tersebut, diterapkan metode imputasi, yaitu mengisi nilai-nilai kosong tersebut dengan nilai median setiap kolom. Strategi median dipilih karena lebih robust atau tahan terhadap nilai-nilai ekstrem (*outlier*) dibandingkan dengan nilai rata-rata (*mean*) [13].
b. Transformasi Data: Algoritma K-Means sangat sensitif terhadap rentang nilai setiap atribut. Atribut dengan rentang nilai yang besar dapat mendominasi proses perhitungan jarak. Oleh karena itu, Proses normalisasi dijalankan menggunakan pendekatan *Min-Max Scaler*. Pendekatan ini mengubah setiap nilai data menjadi rentang seragam antara 0 dan 1, sehingga setiap atribut memiliki bobot yang sama dalam proses klasterisasi [12].

3.3. Proses Klasterisasi dengan K-Means

Tahap ini merupakan inti dari proses *data mining* dalam penelitian ini.

a. Penentuan Jumlah Klaster (k) Optimal: Salah satu prasyarat utama di dalam pendekatan K-Means, salah satu langkah penting adalah menetapkan berapa banyak kelompok (k) yang akan terbentuk. Jumlah k optimal dalam penelitian ini ditentukan dengan menggunakan *Elbow Method*. Metode ini bekerja dengan cara menjalankan algoritma K-Means untuk berbagai nilai k (dalam penelitian ini diujikan mulai dari k=2 hingga k=10) lakukan perhitungan nilai *Sum of Squared Errors* (SSE) atau *inertia* untuk setiap nilai k yang ditentukan. Nilai k optimal dipilih pada titik di mana pengurangan nilai SSE mulai berkurang secara mencolok, sehingga membentuk visualisasi seperti "siku" pada grafik [10].
b. Pemodelan K-Means: Setelah jumlah k optimal diperoleh, maka dibangun model K-Means dengan menggunakan nilai k tersebut. Algoritma kemudian dijalankan pada data *pre-processing* untuk mengategorikan setiap informasi film ke dalam *cluster* yang tepat berdasarkan kedekatannya dengan *centroid cluster*.

3.4. Evaluasi Model Klasterisasi

Evaluasi dilakukan untuk mengukur kualitas struktur *cluster* yang telah terbentuk. Karena klasterisasi merupakan metode *unsupervised learning*, maka evaluasi dilakukan dengan menggunakan metrik validasi internal. Dua metrik utama yang digunakan adalah:

a. *Silhouette Score*: Ukuran ini menilai seberapa optimal suatu objek berada dalam *cluster* yang dituju dibandingkan dengan *cluster-cluster* lainnya. Formula untuk menghitung Silhouette Score bagi satu titik data i dapat dinyatakan dalam rumus berikut:

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)} \quad (1)$$

Di mana a(i) merujuk pada nilai rata-rata jarak dari titik i ke semua titik lainnya dalam *cluster* yang sama, sementara b(i) mencerminkan rata-rata jarak dari titik i ke semua titik dalam *cluster* terdekat berikutnya. Nilai *Silhouette Score* secara keseluruhan merupakan rata-rata dari s(i) untuk setiap titik data, dengan rentang nilai yang bervariasi dari -1 hingga +1. Nilai yang mendekati +1 menunjukkan bahwa data tersebut sangat cocok dengan *cluster* nya dan terpisah dengan baik dari *cluster* yang lain [14].

b. *Davies-Bouldin Index* (DBI): Ukuran ini menilai perbandingan kesamaan rata-rata antara setiap *cluster* dan *cluster* yang memiliki kemiripan tertinggi dengannya. Dalam hal ini, kesamaan diukur sebagai perbandingan antara jarak dalam kelompok intra-*cluster* dan jarak antar kelompok inter-*cluster*. Rumus DBI adalah:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i}\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)}\right) \quad (2)$$

Di dalam konteks ini, k merujuk pada total jumlah *cluster*, σ melambangkan rata-rata jarak seluruh data di dalam satu *cluster* dengan *centroid* nya, sementara d(c_i, c_j) menunjukkan jarak antara *centroid* dari *cluster* i dan j. Semakin kecil nilai DBI dekat dengan 0, semakin baik hasil klasterisasi yang diperoleh, di mana *cluster* yang dihasilkan menjadi lebih kompak dan saling terpisah dengan jelas satu sama lain [8].

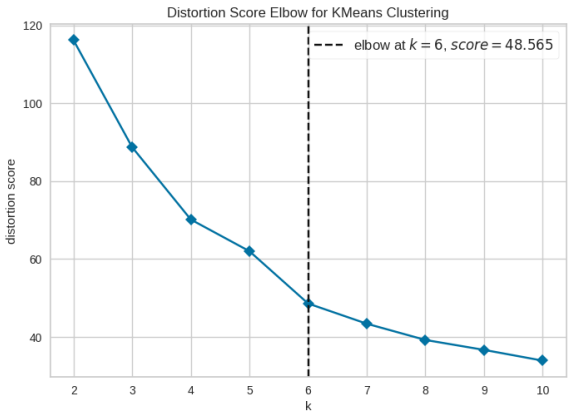
4. HASIL DAN PEMBAHASAN

Bagian ini akan memaparkan hasil proses klasterisasi data film Amazon Prime Video menggunakan algoritma K-Means. Pembahasan meliputi hasil penentuan jumlah *cluster* yang optimal, evaluasi model yang terbentuk, serta analisis mendalam terhadap profil dan karakteristik masing-masing *cluster* yang berhasil diidentifikasi.

4.1. Hasil Penentuan Jumlah Klaster (k)

Selaras dengan pendekatan yang telah dipaparkan dalam bab 3, penentuan jumlah klaster (k) yang optimal dilakukan dengan menggunakan *Elbow Method*. Metode ini bertujuan untuk mencari titik dimana penambahan jumlah *cluster* tidak lagi memberikan penurunan *Sum of Squared Errors* (SSE) atau nilai *inertia* yang signifikan. Analisis menggunakan *Method Elbow* dapat diamati pada Gambar 2. yang tertera di bawah ini.

Proses penentuan nilai k dalam grafik tersebut melibatkan evaluasi terhadap laju penurunan *Distortion Score* atau *inertia*. *Inertia* sendiri merupakan ukuran yang menunjukkan seberapa besar jarak kuadrat rata-rata dari setiap sampel ke *centroid* dari *cluster* yang bersangkutan. Dalam Gambar 2. dapat dilihat bahwa saat nilai k bertambah dari 2 ke 3, dan seterusnya, terjadi penurunan yang tajam pada nilai *inertia*. Hal ini menunjukkan bahwa penambahan *cluster* secara substansial meningkatkan kepadatan dalam *cluster* tersebut. Namun, setelah mencapai k=6, laju penurunan *inertia* menjadi lebih datar dan tidak mengalami perubahan yang signifikan. Titik di mana k=6 ini dikenal sebagai "siku" (*elbow*), di mana tercapai keseimbangan optimal antara jumlah *cluster* dan nilai *inertia*. Penambahan lebih banyak *cluster* setelah titik tersebut tidak memberikan peningkatan kualitas model yang proporsional. Dengan mempertimbangkan analisis ini, nilai k=6 dipilih sebagai jumlah *cluster* yang optimal untuk kajian ini [10].



Gambar 2. Hasil Analisis *Elbow Method* untuk Menentukan Jumlah Klaster Optimal

4.2. Hasil Evaluasi Model

Setelah model K-Means dibangun dengan k=6, dilakukan evaluasi untuk mengukur kualitas struktur *cluster* yang terbentuk. Hasil evaluasi menggunakan metrik validasi internal adalah sebagai berikut:

a. Silhouette Score: Model menghasilkan nilai *Silhouette Score* sebesar 0,3375. Nilai ini berada pada rentang positif dan cukup jauh dari angka 0 menunjukkan bahwa semburat struktur *cluster* yang terbentuk tergolong memuaskan. Ini berarti, secara umum, setiap film lebih memiliki kesamaan dengan film-film lain yang berada dalam *cluster* nya ketimbang dengan film-film yang tergolong dalam *cluster* yang berbeda [14].

b. Davies-Bouldin Score: Model menghasilkan nilai *Davies-Bouldin Score* sebesar 0,9418. Nilai DBI yang mendekati 0 dianggap lebih baik. Hasil ini mengindikasikan bahwa *cluster* yang telah terbentuk menunjukkan tingkat pemisahan yang cukup baik satu dengan yang lainnya. [8].

Kedua hasil evaluasi ini secara kuantitatif menunjukkan bahwa model klasterisasi yang dibangun memiliki kualitas yang valid dan dapat diandalkan untuk analisis lebih lanjut.

4.3. Profil dan Analisis Klaster

Proses klasterisasi berhasil mengelompokkan 2.108 data film ke dalam 6 *cluster* yang berbeda. Untuk mengidentifikasi ciri khas dari setiap *cluster*,

analisis profil dilaksanakan dengan cara menghitung nilai rata-rata dari setiap atribut untuk setiap *cluster* yang ada. Temuan ini dapat dilihat pada Tabel 1. yang tersaji di bawah ini.

Tabel 1. Profil Rata-rata Karakteristik per Klaster

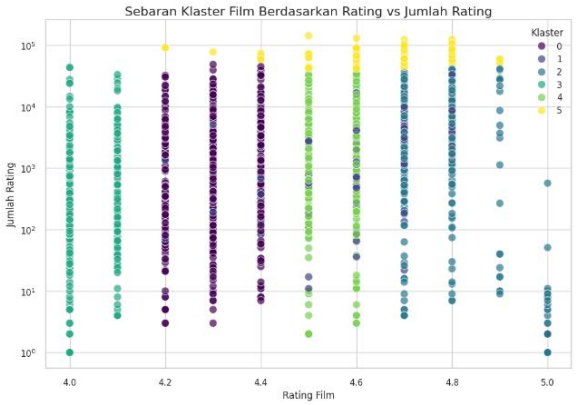
Cluster	Movie Rating	No of Ratings	Release Year	Price
0	4.31	3827	2014	3.94
1	4.55	5385	2012	4.48
2	4.05	2156	2014	3.88
3	4.62	4086	1978	3.78
4	4.77	9207	2010	4.54
5	4.67	69204	2015	6.68

Berdasarkan Tabel 1. setiap *cluster* dapat diartikan sebagai segmen film yang unik:

a. Klaster 5: Film Laris yang Sangat Populer & Premium
   Klaster ini adalah segmen yang paling menonjol. Segmen ini dicirikan oleh tingkat popularitas yang sangat tinggi (No_of_Ratings) (rata-rata 69.204), jauh melampaui *cluster* lainnya. Film-film dalam segmen ini memiliki peringkat yang sangat baik (4,67) dan merupakan film modern (rata-rata tahun 2015) dengan harga tertinggi (rata-rata $6,68). Klaster ini mewakili film-film yang sangat populer dan menjadi daya tarik utama platform ini.
b. Klaster 4: Film Modern Berkualitas Tinggi
   Segmen ini memiliki peringkat rata-rata tertinggi di antara semua *cluster* (4,77). Popularitasnya juga sangat tinggi (rata-rata 9.207), menjadikannya *cluster* terpopuler kedua. Film-film ini merupakan produksi modern (rata-rata tahun 2010) yang sangat diapresiasi oleh penonton karena kualitasnya.
c. Klaster 3: Film Klasik & Lama
   Ciri utama klaster ini adalah tahun rilisnya yang sangat lama, dengan rata-rata tahun 1978. Meskipun sudah tua, film-film ini tetap memiliki rating yang sangat baik (4,62) dengan harga yang terjangkau. Popularitasnya di era modern tidak terlalu tinggi, yang menunjukkan bahwa segmen ini berisi film-film klasik berkualitas yang dinikmati oleh penonton tertentu.
d. Klaster 2: Film Niche atau Kurang Populer
   Klaster ini dicirikan oleh metrik kinerja terendah. Rating rata-ratanya adalah yang terendah (4,05) dan popularitasnya juga terendah (rata-rata 2,156). Ini adalah segmen film yang kemungkinan besar gagal menarik banyak penonton atau tidak menerima ulasan yang baik.
e. Klaster 1 dan Klaster 0: Film Modern Standar dan Populer
   Kedua *cluster* ini mewakili sebagian besar film modern. Klaster 0 adalah film modern "standar" (rata-rata tahun 2014) dengan rating yang wajar (4,31) dan popularitas sedang. Klaster 4 memiliki profil yang sedikit lebih baik, dengan rating yang lebih tinggi (4,55), popularitas yang lebih tinggi (rata-rata 5,385), dan tahun rilis yang sedikit lebih tua (rata-rata tahun 2012). Keduanya merupakan tulang punggung katalog film modern.

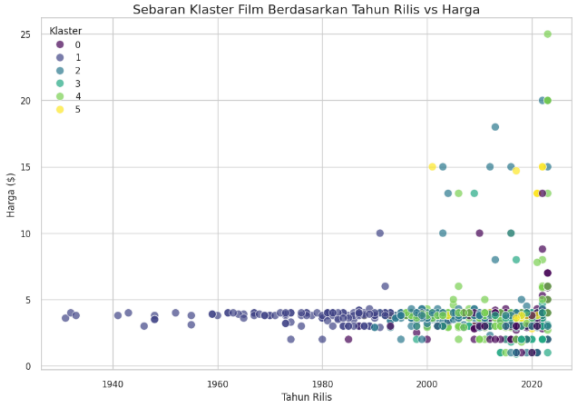
4.4. Pembahasan

Secara keseluruhan, hasil dari proses klasterisasi menunjukkan bahwa algoritma K-Means berhasil dalam menciptakan segmentasi yang signifikan pada katalog film di Amazon Prime Video. Enam segmen yang terbentuk tidak hanya berbeda dalam aspek statistik, tetapi juga menampilkan identitas pasar yang sangat jelas. Penemuan ini menunjukkan bahwa pasar film di platform *streaming* tidaklah seragam, melainkan terdistribusi ke dalam berbagai rongga yang berbeda. Visualisasi klasterisasi yang ditampilkan pada Gambar 3. dan Gambar 4. dibawah semakin mendukung analisis ini. Berikut dibawah ini adalah visualisasi dari hasil klasterisasinya:



Gambar 3. Sebaran Klaster Berdasarkan Rating dan Popularitas

Pada Gambar 3. terlihat dengan jelas pemisahan Klaster 5 yang dikenal sebagai 'Blockbuster Sangat Populer & Premium' dari kelompok lainnya, dikarenakan nilai No_of_Ratings yang sangat tinggi. Sementara itu, Klaster 3 yang disebut 'Film Niche atau Kurang Populer' tampak terkonsentrasi di bagian bawah, menandakan peringkat dan popularitas yang relatif rendah.



Gambar 4. Sebaran Klaster Berdasarkan Tahun Rilis dan Harga

Di sisi lain, pada Gambar 4. terlihat dengan jelas pemisahan Klaster 1 yang teridentifikasi sebagai "Film Klasik & Jadul" di bagian kiri grafik, menguatkan identitasnya sebagai segmen dari film-film lawas. Lebih jauh, pengidentifikasian Klaster 5 sejalan dengan penelitian yang dilakukan oleh Suarna (2023) dan Fitrianti (2023), yang juga mengungkap adanya segmen film populer pada platform Netflix. Namun, studi ini memberikan dimensi baru dengan menyoroti fakta bahwa segmen yang populer di Amazon Prime menunjukkan karakteristik harga rata-rata yang jauh lebih tinggi, mengindikasikan adanya kemungkinan strategi monetisasi yang berbeda. Dari sudut pandang praktis, segmentasi ini membawa implikasi strategis yang sangat signifikan. Amazon Prime Video memiliki kesempatan untuk merancang pendekatan yang berbeda untuk setiap segmennya. Sebagai contoh, film dari Klaster 5 bisa menjadi inti dari kampanye untuk menarik pengguna baru. Film yang termasuk dalam Klaster 2 ("Film Modern Berkualitas Tinggi"), yang menonjol dalam hal rating, dapat secara aktif dipromosikan kepada pengguna yang menunjukkan ketertarikan terhadap film berkualitas tinggi, sementara film dari Klaster 1 dapat ditujukan kepada penonton yang menyukai sinema klasik [6], [7]. Temuan ini akan dirangkum lebih lanjut dalam bagian kesimpulan penelitian.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, disimpulkan bahwa penerapan algoritma K-Means mampu dengan sukses mengelompokkan sejumlah 2.108 film yang tersedia di platform Amazon Prime Video. Melalui *Elbow Method*, ditetapkan bahwa jumlah *cluster* yang paling optimal adalah 6, di mana kualitas model tersebut divalidasi dengan nilai *Silhouette Score* yang mencapai 0.3375 serta *Davies-Bouldin Score* sebesar 0.9418. Setiap *cluster* yang terbentuk mencerminkan segmen pasar yang unik, seperti segmen "Blockbuster Sangat Populer & Premium" (Klaster 5), yang memiliki rata-rata popularitas 69.204 rating, dan segmen "Film Modern Berkualitas Tinggi" (Klaster 2), yang menunjukkan rata-rata rating tertinggi sebesar 4.77. Temuan yang bersifat kuantitatif dan kualitatif ini semakin menegaskan bahwa K-Means adalah metode yang efektif dalam melakukan segmentasi katalog konten digital, memberikan wawasan berharga untuk pengembangan strategi pemasaran serta sistem rekomendasi. Untuk penelitian selanjutnya, direkomendasikan untuk melakukan analisis perbandingan dengan menggunakan algoritma *clustering* alternatif seperti DBSCAN, serta memperkaya analisis dengan memasukkan atribut kategorikal seperti genre dan sutradara.

DAFTAR PUSTAKA

[1] C. Budihartanti, C. I. Ifaru, A. Zahra, dan M. H. Aenuddin, "Pengelompokkan Film Pada Platform Netflix Menggunakan Metode K-Means Clustering Sebagai Rekomendasi Film," *vol*, vol. 5, hlm. 1392–1402, 2024.

[2] N. Suarna, N. Hidayah, dan W. Prihartono, "PENGELOMPOKAN DATA FILM PADA NETFLIX MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, hlm. 3834–3842, Feb 2024, doi: 10.36040/jati.v7i6.8231.

[3] I. Fitrianti, A. Voutama, dan Y. Umaidah, "Clustering Film Populer pada Aplikasi Netflix dengan Menggunakan Algoritma K-Means dan Metode CRISP-DM," *Jurnal Teknologi Sistem Informasi*, vol. 4, no. 2, hlm. 301–311, 2023.

[4] D. Remawati, H. Wijayanto, Y. R. Wahyu Utami, dan B. D. Raharja, "Pengelompokkan Film Trending di Youtube Menggunakan TF-IDF dan K-Means Clastering," *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 4, no. 1, hlm. 65–74, Jan 2025, doi: 10.53513/jursi.v4i1.10614.

[5] R. Aji Sasmoyo, "SISTEM CLUSTERING DAERAH RAWAN KEMATIAN ANAK DI BAWAH UMUR DI WILAYAH JAWA BARAT MENGGUNAKAN ALGORITMA K-MEANS," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, hlm. 5635–5640, Jun 2024, doi: 10.36040/jati.v8i4.9965.

[6] M. Billah, M. A. Zartesya, dan D. S. Prasvita, "Penerapan Collaborative Filtering, PCA dan K-Means dalam Pembangunan Sistem Rekomendasi Film," dalam *Prosiding Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya*, 2021, hlm. 579–587.

[7] Elsa Vania, Salma Nuraini, dan D. S. Y. Kartika, "PENGGUNAAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN REKOMENDASI FILM INDONESIA," *Prosiding Seminar Nasional Teknologi dan Sistem Informasi*, vol. 2, no. 1, hlm. 207–214, Sep 2022, doi: 10.33005/sitasi.v2i1.299.

[8] I. F. Ashari, R. Banjarnahor, D. R. Farida, S. P. Aisyah, A. P. Dewi, dan N. Humaya, "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies," *Journal of Applied Informatics and Computing*, vol. 6, no. 1, hlm. 07–15, Jul 2022, doi: 10.30871/jaic.v6i1.3485.

[9] P. Parlaungan, F. A. Mustika, dan H. Dhika, "Sistem Rekomendasi Film Menggunakan Metode K-Means Clustering Berbasis Web," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 13, no. 2, 2022.

[10] T. Nugroho, U. Enri, dan I. Maulana, "CLUSTERING MENU MAKANAN BERDASARKAN KEBUTUHAN KALORI HARIAN MENGGUNAKAN ALGORITME K-MEANS," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 4, hlm. 4411–4417, Jun 2024, doi: 10.36040/jati.v8i4.9958.

[11] N. Zalzabila dan R. Prathivi, "Analisis Preferensi Penonton Anime berbasiskan Genre Film menggunakan Metode K-Means," *Kesatria : Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, vol. 6, no. 1, hlm. 235–241, Jan 2025, doi: 10.30645/kesatria.v6i1.565.

[12] Z. Y. M. Gumiwang, A. F. Hamidy, dan E. Y. Puspaningrum, "Analisis Data Mining Untuk Clustering Data Film Dengan Menggunakan Algoritma K-Means," *Jurnal Manajamen Informatika Jayakarta*, vol. 3, no. 1, hlm. 52–56, 2023.

[13] M. Muhajir dan A. A. P. Sari, "Implementasi Metode Improved K-Means dengan Algoritma Dbscan untuk Pengelompokan Film," *Unisda Journal of Mathematics and Computer Science (UJMC)*, vol. 6, no. 01, hlm. 1–8, Jun 2020, doi: 10.52166/ujmc.v6i01.1923.

[14] A. Fatlun, S. S. Hilabi, B. Priyatna, dan A. L. Hananto, "Penggunaan Algoritma K-Means Clustering CRISP-DM Dalam Mengelompokkan Drama Korea Sebagai Rekomendasi Film," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 12, no. 1, Mar 2025, doi: 10.35957/jatisi.v12i1.11347