

CLUSTERING DATA FILM DAN SERIES NETFLIX MENGUNAKAN METODE K-MEANS DAN AGGLOMERATIVE HIERARCHICAL CLUSTERING

Nathaniel Andrew

Program Studi Teknik Informatika, Universitas Tarumanagara
Jl. Letjen S.Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
e-mail: *nathaniel.535200054@stu.untar.ac.id*

ABSTRAK

Film dan series di Netflix dapat dicluster berdasarkan genre, negara asal, dan tema. *Cluster* berdasarkan genre dapat membantu pengguna menemukan film atau series yang sesuai dengan minat mereka. *Cluster* berdasarkan negara asal dapat membantu pengguna menemukan karya-karya dari berbagai negara. *Cluster* berdasarkan tema dapat membantu pengguna menemukan film atau series yang mengangkat tema tertentu. *Clustering* film dan series di Netflix dapat membantu pengguna menemukan tontonan yang sesuai dengan minat dan kebutuhan mereka. Hal ini dapat meningkatkan kepuasan pengguna dengan layanan Netflix.

Kata kunci: *Clustering, Netflix, K-Means, Agglomerative.*

ABSTRACT

Films and series on Netflix can be clustered by genre, country of origin and theme. Clusters based on genre can help users find films or series that suit their interests. Clusters based on country of origin can help users find works from various countries. Theme-based clusters can help users find films or series that cover certain themes. Clustering films and series on Netflix can help users find shows that suit their interests and needs. This can increase user satisfaction with Netflix services.

Keywords: *Clustering, Netflix, K-Means, Agglomerative.*

1 PENDAHULUAN

Netflix memiliki perpustakaan film dan series yang sangat besar, yang dapat menjadi tantangan bagi pengguna untuk menemukan konten yang mereka sukai. *Clustering* adalah teknik pemrosesan data yang dapat digunakan untuk mengelompokkan konten berdasarkan kemiripan. Dengan mengelompokkan konten, Netflix dapat membuat lebih mudah bagi pengguna untuk menemukan konten yang mereka minati.

Dalam penelitian ini, kami mengusulkan pendekatan untuk *clustering* film dan series Netflix menggunakan Python. Pendekatan kami menggunakan metode *agglomerative hierarchical clustering* untuk mengelompokkan konten berdasarkan kemiripan topik. Metode *agglomerative hierarchical clustering* adalah metode *clustering* yang bekerja dengan menggabungkan dua *cluster* yang paling mirip pada setiap Langkah [1]. Proses ini terus berlanjut sampai semua *cluster* telah digabungkan menjadi satu *cluster*.

Untuk mengukur kemiripan topik antara dua dokumen, kami menggunakan teknik TF-IDF [2]. TF-IDF adalah ukuran statistik yang digunakan untuk mengukur pentingnya suatu istilah terhadap sebuah dokumen dalam sebuah kumpulan atau korpus. TF-IDF dihitung dengan cara membagi jumlah kemunculan suatu istilah dalam sebuah dokumen dengan jumlah total kata dalam dokumen tersebut. Kemudian, hasil ini dibagi dengan jumlah dokumen dalam kumpulan atau korpus yang mengandung istilah tersebut [3]. Setelah kami memiliki data, kami perlu memprosesnya untuk menghilangkan *stopword* dan menormalkan kata-kata. Kami juga perlu menghitung TF-IDF untuk setiap dokumen.

Setelah kami memproses data, kami dapat menggunakan metode *agglomerative hierarchical clustering* untuk mengelompokkan konten. Dalam penelitian kami, kami menemukan bahwa metode kami dapat dengan efektif mengelompokkan film dan series Netflix berdasarkan kemiripan topik [4]. Kami juga menemukan bahwa metode kami dapat digunakan untuk menghasilkan *cluster* yang berukuran wajar.

2 TINJAUAN LITERATUR

Penelitian ini bertujuan untuk mendalami pemahaman tentang jenis-jenis film dan seri Netflix melalui penerapan teknik clustering pada dataset yang mencakup berbagai atribut, khususnya fokus pada karakteristik tertentu seperti 'Genre' dan 'Rating'. Pertama, penelitian ini bertujuan untuk mengidentifikasi pola dan hubungan yang mungkin ada di antara data, memungkinkan kita untuk merinci struktur inherent dalam dataset film dan seri Netflix. Selanjutnya, penelitian ini akan mengeksplorasi potensi klasifikasi atau kelompok berdasarkan atribut-atribut tersebut, menyediakan landasan bagi pemahaman lebih mendalam tentang variasi dalam jenis film dan seri Netflix.

Dalam melakukan pra-pemrosesan data pada dataset film dan seri Netflix, beberapa langkah krusial telah diimplementasikan. Pertama, dilakukan pemilihan fitur, di mana fitur-fitur yang paling relevan untuk analisis *clustering* dipilih, dengan fokus utama pada atribut seperti 'Genre' dan 'Rating'. Langkah kedua melibatkan standarisasi data untuk mengatasi potensi masalah skala antar fitur. Oleh karena itu, dilakukan transformasi data menggunakan StandardScaler dari *scikit-learn*, memastikan bahwa distribusi fitur-fitur tersebut seragam. Setelah itu, dilakukan visualisasi data menggunakan *scatter plot* untuk memberikan gambaran jelas tentang distribusi sebelum proses *clustering*. Langkah keempat melibatkan penanganan missing values dan pembersihan data, dengan mengidentifikasi dan menangani nilai-nilai yang hilang serta *outliers* yang mungkin mempengaruhi hasil clustering. Langkah terakhir, namun tidak kalah pentingnya, adalah ekstraksi fitur jika diperlukan untuk meminimalisir kesalahan data atau mengekstrak informasi yang lebih relevan.

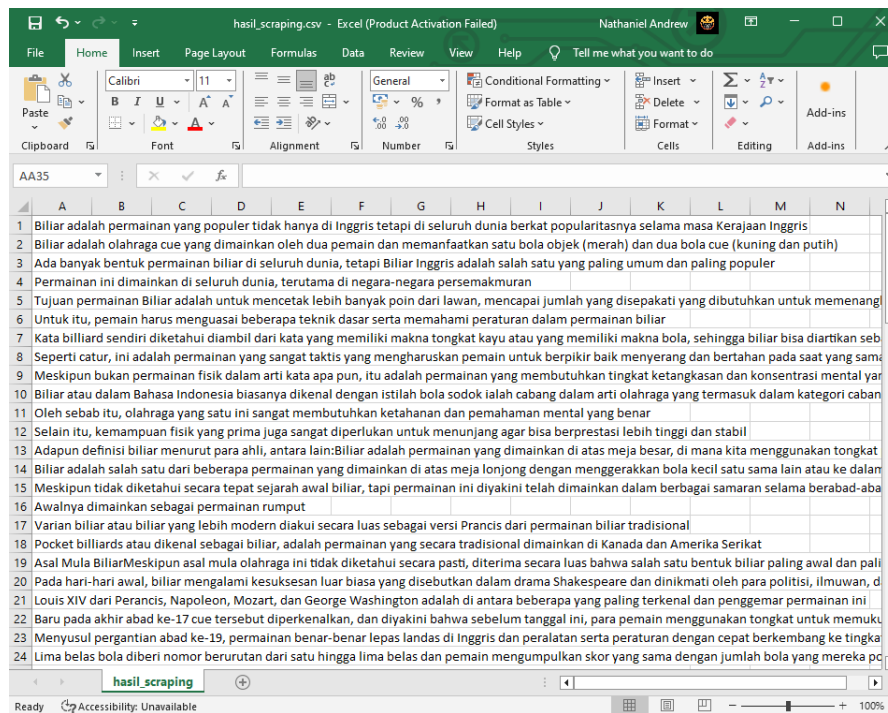
Dalam konteks ini, tujuan lainnya adalah membandingkan efektivitas *algoritma clustering*, baik *K-means* maupun *Agglomerative* untuk membentuk kelompok-kelompok yang bermakna dalam konteks film dan seri Netflix. Evaluasi menggunakan metrik seperti *silhouette score* diharapkan memberikan dasar perhitungan skor yang terbaik untuk mengukur kualitas pembentukan kelompok. Setelah proses pra-pemrosesan data selesai, langkah selanjutnya adalah menerapkan algoritma *clustering*, yaitu *K-means* dan *Agglomerative*, untuk mengelompokkan film dan seri Netflix berdasarkan atribut yang telah dipilih. *K-means* akan berusaha mengelompokkan data menjadi k kelompok yang memiliki pusat kluster tertentu, sedangkan *Agglomerative* akan membentuk hirarki kluster dengan menggabungkan kluster secara bertahap. Selain itu, dilakukan analisis visual terhadap hasil *clustering* dengan menggunakan *plot* atau visualisasi lainnya untuk memberikan pemahaman yang lebih mendalam tentang kelompok-kelompok yang terbentuk. Ini dapat membantu mengidentifikasi pola atau tren yang mungkin tersembunyi dalam data.

3 METODE PENELITIAN

Berdasarkan hasil studi dan analisis yang telah dilakukan, diperlukan sebuah rancangan sistem yang mencakup metode serta sumber data untuk implementasi lebih lanjut. Rancangan ini akan diuraikan pada sub bab berikut, menyajikan detail tentang bagaimana sistem ini akan dibentuk.

3.1 Dataset

Dalam penelitian *clustering* yang dilakukan, dataset yang digunakan diambil melalui *website* www.kaggle.com yang berupa dataset berisikan berbagai jenis judul film dan series yang terdapat pada netflix.



Gambar 1. Dataset Netflix Berbentuk CSV

	type	title	director	cast	country	date_added	release_year	rating	duration
show_id									
s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons
s2	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min
s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min
s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 min
s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13	123 min

Gambar 2. Atribut dan variable dari dataset

Dapat dilihat pada Gambar 1 dan 2 merupakan contoh dataset netflix yang berisikan atribut dan variable dari data yang digunakan yaitu dataset netflix yang terdapat jenis film, sutradara, pemeran, dan lainnya.

3.1 Pre - processing

Pre-processing adalah langkah awal yang penting setelah mengumpulkan data mentah atau data yang belum diolah. Tujuannya adalah membersihkan data, mengelompokkan data, dan melakukan persiapan agar data lebih mudah dianalisis dan diproses oleh algoritma perhitungan.

1. *Tokenizing*: Ini adalah proses memecah teks menjadi bagian-bagian yang lebih kecil yang disebut sebagai token. Token bisa berupa kata, simbol, atau angka yang memiliki makna tersendiri..
2. *Case Folding*: Tahap ini mengubah huruf dalam teks menjadi huruf besar atau kecil sesuai dengan kebutuhan.
3. *Lemmatization*: Proses mencari kata dasar dari sebuah kata bentukan. Kata bentukan adalah kata yang dibentuk dari kata dasar melalui proses morfologis, seperti pengimbuhan, pemajemukan, atau pengulangan..
4. *Punctuation Removal*: Tahap ini melibatkan penghapusan karakter tanda baca yang tidak relevan dalam teks, seperti titik, koma, tanda seru, atau tanda tanya.
5. *TF-IDF*: Kemunculan suatu istilah dalam sebuah dokumen. TF dihitung dengan cara membagi jumlah kemunculan suatu istilah dalam sebuah dokumen dengan jumlah total kata dalam dokumen tersebut.

```
[ ] import nltk
    from nltk.corpus import stopwords
    sw = stopwords.words('english')
    np.array(sw)

array(['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you',
      'you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself',
      'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her',
      'hers', 'herself', 'it', "it's", 'its', 'itself', 'they', 'them',
      'their', 'theirs', 'themselves', 'what', 'which', 'who', 'whom',
      'this', 'that', "that'll", 'these', 'those', 'am', 'is', 'are',
      'was', 'were', 'be', 'been', 'being', 'have', 'has', 'had',
      'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'and',
      'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of', 'at',
      'by', 'for', 'with', 'about', 'against', 'between', 'into',
      'through', 'during', 'before', 'after', 'above', 'below', 'to',
      'from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under',
      'again', 'further', 'then', 'once', 'here', 'there', 'when',
      'where', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more',
      'most', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own',
      'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will',
      'just', 'don', "don't", 'should', "should've", 'now', 'd', 'll',
      'm', 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn',
      "couldn't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't",
      'hasn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma',
      'mightn', "mightn't", 'mustn', "mustn't", 'needn', "needn't",
      'shan', "shan't", 'shouldn', "shouldn't", 'wasn', "wasn't",
      'weren', "weren't", 'won', "won't", 'wouldn', "wouldn't"],
      dtype='<U10')

```

Gambar 3. Implementasi Stopwords

```
[ ] def remove_punctuation(text):
    '''a function for removing punctuation'''
    translator = str.maketrans('', '', string.punctuation)
    return text.translate(translator)

[ ] df1['clustering_attributes'] = df1['clustering_attributes'].apply(remove_punctuation)

df1['clustering_attributes'][40]

'muharrem gulmez erdem yener ayhan tas emin olcay muharrem gulmez elif nur kerkuk tark
ity save big developer comes town close open new mall'
```

Gambar 4. Implementasi Punctuation Removal

```
def lemmatize_verbs(words):
    """Lemmatize verbs in list of tokenized words"""
    lemmatizer = WordNetLemmatizer()
    lemmas = []
    for word in words:
        lemma = lemmatizer.lemmatize(word, pos='v')
        lemmas.append(lemma)
    return lemmas

df1['clustering_attributes'] = lemmatize_verbs(df1['clustering_attributes'])

df1['clustering_attributes'][40]
```

Gambar 5. Implementasi lemmatize

```
[ ] tokenizer = TweetTokenizer()

[ ] df1['clustering_attributes'] = df1['clustering_attributes'].apply(lambda x: tokenizer.tokenize(x))

clustering_data = df1['clustering_attributes']

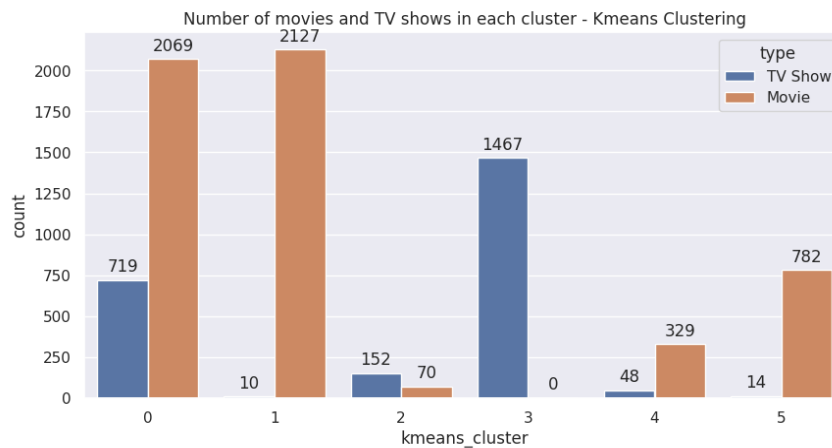
def identity_tokenizer(text):
    return text

tfidf = TfidfVectorizer(tokenizer=identity_tokenizer, stop_words='english', lowercase=False, max_features = 20000)
X = tfidf.fit_transform(clustering_data)
```

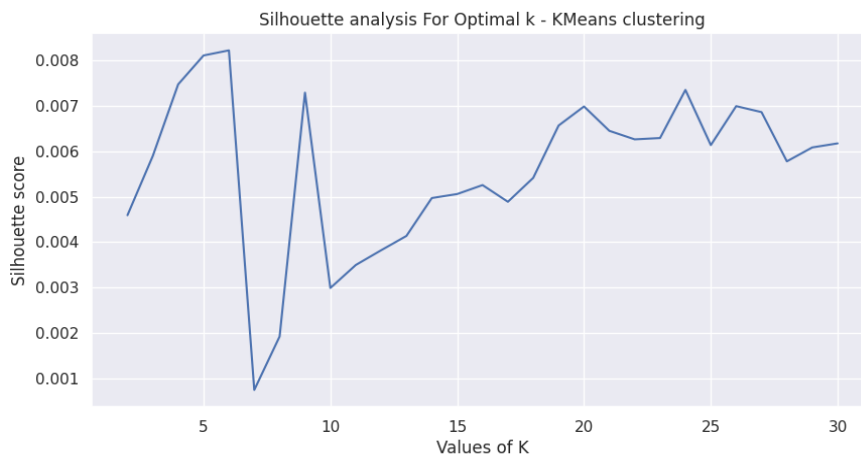
Gambar 6. Implementasi Tokenize dan TF-IDF

3.2 K-Means

Metode *K-Means* adalah suatu algoritma *clustering* yang digunakan untuk mengelompokkan data ke dalam kluster atau kelompok yang serupa. Tujuannya adalah untuk meminimalkan varian intra-kluster dan memaksimalkan varian antara kluster. Prosesnya melibatkan langkah-langkah iteratif untuk menentukan pusat kluster dan mengelompokkan data berdasarkan kedekatan dengan pusat kluster tersebut. Metode *K-Means* mengoptimalkan fungsi tujuan yang mencakup total kuadrat jarak antara setiap data dengan pusat kluster. Proses ini menghasilkan pembentukan kelompok data yang memiliki kemiripan tinggi dalam satu kluster dan perbedaan yang signifikan antara kluster [7]. keberhasilan metode *K-Means* sangat dipengaruhi oleh pemilihan jumlah kluster yang tepat dan inisialisasi pusat kluster yang baik. Evaluasi kualitas *clustering* dapat dilakukan dengan menggunakan metrik seperti *silhouette score* atau validasi internal dan eksternal lainnya.



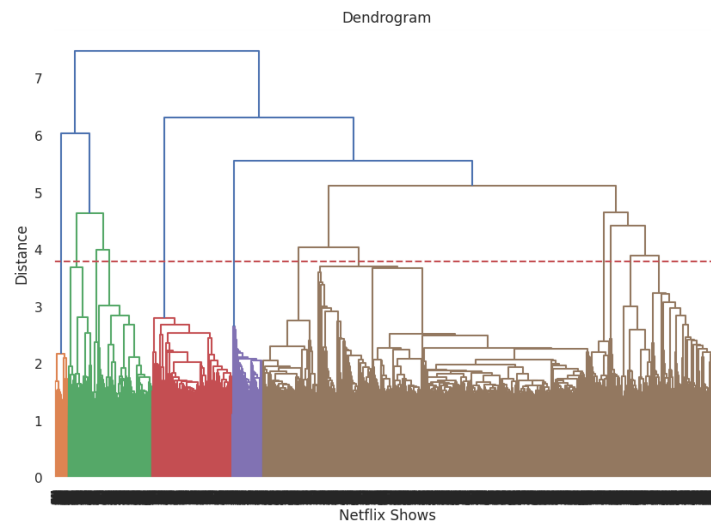
Gambar 6. Hasil Cluster Dari Metode K-Means



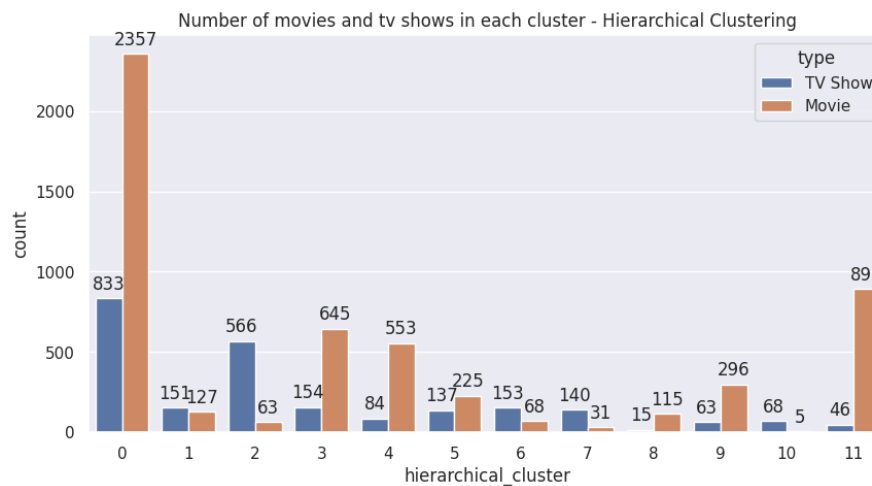
Gambar 6. Silhouette Score Dari Metode K-Means

3.3 Agglomerative Hierarchical Clustering

Metode *Agglomerative Hierarchical Clustering* (AHC) adalah suatu algoritma *clustering* yang bekerja dengan cara mengelompokkan data secara hierarkis. Prosesnya dimulai dengan setiap data dianggap sebagai kluster individual dan secara iteratif menggabungkan kluster-kluster yang memiliki kedekatan atau kemiripan tertinggi. Proses penggabungan ini terus berlanjut hingga semua data tergabung dalam satu kluster utama [8]. Metode AHC menghasilkan struktur hirarkis atau dendrogram yang merepresentasikan hubungan hierarkis antar kluster. Pada dendrogram, tinggi cabang atau garis horizontal mengindikasikan tingkat kemiripan atau kedekatan antar kluster. Pemotongan dendrogram pada suatu tingkat tertentu dapat menghasilkan sejumlah kluster yang diinginkan [9].



Gambar 7. Dendrogram Jarak Antar *Cluster* Metode AHC



Gambar 8. Hasil *Cluster* Dari Metode AHC

3.3 Skema Eksperimen

Skema eksperimen yang dilakukan untuk clustering pada dataset film dan seri Netflix melibatkan beberapa tahap utama. Pertama, akan dilakukan pra-pemrosesan data, di mana fitur-fitur yang paling relevan untuk analisis clustering akan dipilih, dan data akan disesuaikan melalui standarisasi menggunakan *Standard Scaler*. Visualisasi data awal akan dilakukan melalui *scatter plot* untuk memahami distribusi sebelum proses *clustering*.

Langkah berikutnya adalah menerapkan dua metode clustering, yaitu *K-Means* dan *Agglomerative Hierarchical Clustering* (AHC), untuk membandingkan pendekatan eksklusif dan hirarkis dalam pembentukan kelompok. *K-Means* akan digunakan dengan jumlah klaster (*n-clusters*) sebanyak 2, sedangkan AHC akan menghasilkan struktur hirarkis berdasarkan tingkat kemiripan antar data.

Setelah klaster-klaster dibentuk, evaluasi kualitas *clustering* akan dilakukan menggunakan metrik seperti *silhouette score matrix* untuk kedua metode. *Silhouette score* akan memberikan informasi tentang sejauh mana kelompok-kelompok tersebut terpisah dan sejauh mana data dalam klaster homogen. Evaluasi ini akan memberikan pemahaman lebih mendalam tentang performa dan karakteristik masing-masing metode *clustering* dalam konteks dataset film dan seri Netflix.

Penting untuk dicatat bahwa eksperimen ini tidak hanya bertujuan untuk mengelompokkan data, tetapi juga untuk memberikan wawasan baru terkait dengan preferensi pengguna dan karakteristik film dan seri. Hasil dari eksperimen ini diharapkan dapat memberikan kontribusi pada pemahaman industri hiburan serta memberikan dasar untuk analisis lanjutan terkait dengan pengembangan konten dan rekomendasi kepada pengguna.

4 HASIL DAN PEMBAHASAN

Hasil dari percobaan dan pembahasan dari eksperimen clustering pada dataset film dan seri Netflix memberikan wawasan yang berharga terkait dengan struktur internal dan potensi klasifikasi berdasarkan karakteristik tertentu. Penggunaan *K-Means* menghasilkan pembagian data ke dalam dua kluster eksklusif, menunjukkan kemampuan *K-Means* untuk mengidentifikasi kelompok yang jelas dan terpisah. Sebaliknya, *Agglomerative Hierarchical Clustering* (AHC) memberikan gambaran yang lebih hierarkis, dengan pembentukan struktur dendrogram yang mencerminkan tingkat kemiripan antar film dan seri.

Visualisasi hasil clustering menunjukkan bahwa *K-Means* dapat menghasilkan pemisahan yang sangat jelas antara dua kelompok, sedangkan AHC menggambarkan hubungan hierarkis antar kelompok dengan tingkat tumpang tindih yang lebih signifikan. Evaluasi menggunakan *silhouette score* mendukung temuan ini, di mana skor siluet *K-Means* mendekati 1, menandakan bahwa kluster-kluster yang terbentuk memiliki pemisahan yang sangat baik.

Temuan ini menyiratkan bahwa, dalam konteks dataset film dan seri Netflix, pendekatan eksklusif seperti *K-Means* sangat efektif untuk mengidentifikasi pola dan kecenderungan yang jelas dalam preferensi pengguna. Sementara itu, AHC memberikan wawasan tambahan terkait dengan struktur hierarkis dan kemungkinan tumpang tindih antar kelompok.

Penting untuk dicatat bahwa pemilihan metode *clustering* dapat bergantung pada tujuan analisis dan jenis informasi yang diinginkan. Temuan ini memberikan dasar untuk analisis lebih lanjut terkait dengan segmentasi pengguna, rekomendasi konten, dan strategi penawaran dalam industri hiburan digital.

5 KESIMPULAN

Dalam Kesimpulan dari eksperimen *clustering* pada dataset film dan seri Netflix dengan menggunakan metode *K-Means* dan *Agglomerative Hierarchical Clustering* (AHC) memberikan wawasan yang mendalam terkait dengan struktur internal dan potensi klasifikasi berdasarkan karakteristik tertentu. Analisis kedua metode clustering memberikan perspektif yang berbeda, memungkinkan pemahaman yang holistik terkait dengan preferensi pengguna dan karakteristik film dan seri.

Penggunaan *K-Means*, dengan mendekati nilai *silhouette score* sebesar 1, menunjukkan bahwa metode ini sangat efektif dalam membentuk dua kelompok eksklusif dengan pemisahan yang sangat baik antara kelompok tersebut. Hasil ini memberikan indikasi bahwa *K-Means* cocok digunakan untuk mengidentifikasi pola dan kecenderungan yang jelas dalam dataset film dan seri Netflix. Dengan visualisasi yang jelas dan pemisahan yang kuat antar kelompok, *K-Means* dapat menjadi pilihan yang solid untuk analisis segmentasi pengguna atau penentuan strategi konten yang sesuai.

Di sisi lain, penggunaan *Agglomerative Hierarchical Clustering* (AHC) menghasilkan struktur dendrogram yang mencerminkan hubungan hierarkis antar kelompok. Pemisahan antar kelompok tidak sejelas seperti pada *K-Means*, dan tingkat tumpang tindih yang lebih signifikan menandakan adanya kemungkinan variasi dalam preferensi pengguna. Meskipun AHC tidak mencapai nilai

silhouette score sebesar *K-Means*, pendekatan hierarkis ini memberikan perspektif yang berharga tentang kompleksitas struktur data dan mungkin lebih sesuai untuk analisis yang mempertimbangkan hierarki dalam preferensi pengguna.

Kesimpulannya, pemilihan antara *K-Means* dan AHC dalam analisis *clustering* film dan seri Netflix bergantung pada tujuan analisis dan jenis informasi yang diinginkan. *K-Means* memberikan kejelasan dan pemisahan yang baik, sementara AHC menyediakan konteks hierarkis yang relevan. Pemahaman karakteristik kedua metode ini dapat digunakan sebagai dasar untuk pengambilan keputusan yang lebih baik dalam strategi pemasaran, pengembangan konten, dan segmentasi pengguna dalam industri hiburan digital.

UCAPAN TERIMA KASIH

Saya ingin menyampaikan rasa syukur kepada Yang Maha Esa atas kelancaran dalam penulisan penelitian ini. Terima kasih juga kepada para pembaca yang telah menyisihkan waktu untuk membaca penelitian ini. Semoga hasil kerja keras ini dapat memberikan dampak positif bagi perkembangan dunia perfilman dan masyarakat secara luas.

DAFTAR PUSTAKA

- [1] Burhanuddin, Zainudin. "Penerapan Metode *Agglomerative Hierarchical Clustering* dengan Validator *Silhouette Index*." Skripsi 1.413416016 (2020).
- [2] Dhuhita, W. M. P. (2015). Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita.
- [3] Suhirman, Suhirman, and Hero Wintolo. "*System for Determining Public Health Level Using the Agglomerative Hierarchical Clustering Method*." Compiler 8.1 (2019): 95-104.
- [4] Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan.
- [5] Sulastrri H, Gufroni AI. Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. Jurnal Teknologi dan Sistem Informasi. 2017.
- [6] Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa
- [7] Ningrat, D. R., Di Asih, I. M., & Wuryandari, T. (2016). Analisis cluster dengan algoritma K-Means clustering untuk pengelompokan data obligasi korporasi.
- [8] Wilson, Wilson. "Website Klasifikasi Jurnal Berdasarkan Abstrak Dengan Metode *Agglomerative Hierarchical Clustering*", *Jurnal Ilmiah Core IT: Community Research Information Technology* 11.4 (2023).
- [9] Alfariz, Riki. "Implementasi Metode *Agglomerative Hierarchical Clustering* Untuk Pengelompokkan Jurnal Berdasarkan Abstrak Berbasis Website." *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)* 6.1 (2023): 700-706.
- [10] Setyawan, Dimas Ari, and Chastine Fatichah. "*Enhancement of Decision Tree Method Based on Hierarchical Clustering and Dispersion Ratio*." *JUTI: Jurnal Ilmiah Teknologi Informasi* 18.2 (2020): 179-187.
- [11] Muflihan, Yenni, Heri Retnawati, and Agus Kistian. "Analisis cluster dengan metode hierarki untuk pengelompokan sekolah menengah atas berdasarkan raport mutu sekolah di Kabupaten Nagan Raya." *Measurement In Educational Research* 2.1 (2022): 22-33.
- [12] Adek, Rizal Tjut, Rozzy Kesuma Dinata, and Ananda Ditha. "*Online Newspaper Clustering in Aceh using the Agglomerative Hierarchical Clustering Method*." *International Journal of Engineering, Science and Information Technology* 2.1 (2022): 70-75.
- [13] Ula, Mutammimul, et al. "Implementation of Machine Learning in Determining Nutritional Status using the Complete Linkage *Agglomerative Hierarchical Clustering Method*." *Jurnal Mantik* 5.3 (2021): 1910-1914.
- [14] Nofiar, Andri, and Sarjon Defit. "Penentuan Mutu Kelapa Sawit Menggunakan Metode K-Means Clustering." *Jurnal KomtekInfo* 5.3 (2018): 1-9.
- [15] Anjelita, Mawaddah, et al. "Analisis Metode K-Means pada Kasus Ekspor Barang Perhiasan dan Barang Berharga Berdasarkan Negara Tujuan." *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*.

- Vol. 2. No. 1. 2019.
- [16] Dewi, Sinta Maulina, et al. "Analisa Metode K-Means pada Pengelompokan Kriminalitas Menurut Wilayah." *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*. Vol. 2. No. 1. 2019. *Teknologi Informasi dan Komputer* 6.1 (2023): 700-706.
 - [17] Sari, Riyani Wulan, and Dedy Hartama. "Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi." *Seminar Nasional Sains dan Teknologi Informasi (SENSASI)*. Vol. 1. No. 1. 2018.
 - [18] Sembiring, Muhammad Ardiansyah, Raja Tama Andri Agus, and Mustika Fitri Larasti Sibuea. "Penerapan Metode Algoritma K-Means Clustering Untuk Pemetaan Penyebaran Penyakit Demam Berdarah Dengue (DBD)." *Journal of Science and Social Research* 4.3 (2021): 336-341.
 - [19] Rozaq, Abdul. "Implementation of K-Means and Agglomerative Hierarchical Methods to House Clusterization." *Jurnal Media Informatika Budidarma* 6.2 (2022): 933-942.
 - [20] Azmi, Mardhiatul, et al. "Comparison of the Performance of the K-Means and K-Medoids Algorithms in Grouping Regencies/Cities in Sumatera Based on Poverty Indicators." *UNP Journal of Statistics and Data Science* 1.2 (2023): 59-66.