

Sistem Rekomendasi Film pada Netflix menggunakan metode *Collaborative Filtering*

Farrel Ardannur Deswanto
Fakultas Informatika
Universitas Telkom
Bandung

farrelardes@student.telkomuniversity.ac.id

Felicia Dina Widyasari
Fakultas Informatika
Universitas Telkom
Bandung

feliciadina@student.telkomuniversity.ac.id

Felicia Talitha Fathin
Fakultas Informatika
Universitas Telkom
Bandung

felicialitha@student.telkomuniversity.ac.id

Abstract— Industri film mengalami perubahan akibat kemajuan teknologi digital dan popularitas platform streaming. Salah satu tantangan utama yang dihadapi industri film adalah menarik perhatian penonton yang semakin selektif. Untuk mengatasi hal ini, diperlukan sistem rekomendasi yang menggunakan pendekatan sains data untuk memahami pola perilaku dan preferensi penonton. Penelitian ini bertujuan untuk membangun sebuah sistem rekomendasi film yang memanfaatkan algoritma *collaborative filtering*. Selain itu, untuk meningkatkan efisiensi dalam menemukan film yang sesuai dengan preferensi pengguna serta mengoptimalkan penggunaan algoritma *collaborative filtering*. Dataset yang digunakan adalah dataset Netflix Original Films & IMDB Scores, yang mencakup berbagai informasi mengenai film-film yang diproduksi oleh Netflix, seperti judul, genre, tanggal rilis, durasi, skor IMDB, dan bahasa. Metode yang digunakan adalah *collaborative filtering*, dimana data rating dan preferensi film dari pengguna lain digunakan untuk membuat prediksi mengenai film yang mungkin disukai oleh pengguna. Tahapan proyek yang akan dilakukan mencakup persiapan, pengembangan, peluncuran, serta pemeliharaan dan perbaikan sistem rekomendasi film.

Keywords— Sistem Rekomendasi Film, Collaborative Filtering, Preferensi Pengguna, Efisiensi Waktu

I. PENDAHULUAN

Dengan kemajuan teknologi digital dan popularitas platform streaming, industri film dan hiburan mengalami transformasi yang signifikan. Perubahan ini berdampak pada cara penonton mengonsumsi konten audiovisual dengan akses tak terbatas ke berbagai film. Adanya kemudahan akses ini, muncul tantangan baru bagi industri film dalam mempertahankan dan meningkatkan kepuasan bagi penonton. Salah satu tantangan utama yang dihadapi industri film adalah melibatkan penonton yang semakin selektif. Penonton memiliki akses tak terbatas ke berbagai film melalui platform streaming yang menawarkan kenyamanan dan fleksibilitas.

Untuk mengatasi tantangan ini, industri film harus beradaptasi dengan cara baru dalam menjangkau dan mempertahankan penonton yang semakin selektif. Sistem rekomendasi film menjadi krusial dalam menyelesaikan masalah ini. Sistem ini menggunakan algoritma sains data yang mampu menganalisis pola perilaku penonton dan preferensi film. Dengan memahami pola ini, sistem dapat

digunakan untuk memprediksi preferensi penonton dan memberikan rekomendasi film yang sesuai.

Terdapat beberapa metode yang digunakan dalam membuat sistem rekomendasi, yaitu *content-based filtering*, *collaborative filtering*, dan *hybrid*. *Content-based filtering* memanfaatkan informasi tentang karakteristik atau konten dari item yang ingin direkomendasikan serta informasi profil pengguna. *Collaborative filtering* bekerja dengan membangun database yang memberikan preferensi pengguna terhadap suatu item. *Collaborative filtering* memberikan prediksi rating dan personal rekomendasi yang berdasarkan yang disukai pengguna lain yang mempunyai selera yang sama. Sementara *hybrid* merupakan penggabungan *content-based filtering* dan *collaborative filtering*. Dari ketiga metode diatas, perusahaan-perusahaan *e-commerce* besar seperti Amazon, Netflix, dan Pandora umumnya menggunakan *collaborative filtering* sebagai metode utama dalam sistem rekomendasi. Hal ini karena *collaborative filtering* dapat memanfaatkan "*wisdom of the crowd*", yaitu kecenderungan bahwa preferensi atau perilaku kolektif dari sekelompok besar pengguna dapat memberikan indikasi yang kuat tentang preferensi individual.

Oleh karena itu, pada penelitian ini dibangun sistem rekomendasi dengan menggunakan metode *collaborative filtering*. Penerapan dilakukan pada *website* dengan merekomendasikan film untuk pengguna. Adapun dataset yang digunakan dalam penelitian ini adalah dataset Netflix Original Films & IMDB Scores yang diperoleh dari Kaggle yang berisikan informasi rating untuk film oleh pengguna serta informasi konten item film berupa genre.

II. STUDI LITERATUR

A. Sistem Rekomendasi Film

Sistem rekomendasi film adalah teknologi yang dirancang untuk membantu pengguna menemukan film yang mungkin mereka sukai berdasarkan analisis data historis seperti rating, ulasan, dan perilaku menonton. Metode seperti *Collaborative Filtering* digunakan untuk memprediksi film yang diminati berdasarkan penilaian atau umpan balik dari pengguna lain. Sistem ini memberikan rekomendasi yang personal dan relevan, membantu pengguna menemukan film dari berbagai pilihan yang tersedia.

B. Collaborative Filtering

Collaborative Filtering (CF) adalah metode dalam sistem rekomendasi yang bekerja dengan mengumpulkan dan menyaring opini atau umpan balik pengguna untuk

memprediksi item yang mungkin disukai. Metode ini terbagi menjadi dua pendekatan utama, yaitu *user-based*, yang mencari pengguna lain dengan preferensi serupa, dan *item-based*, yang mencari item lain yang mirip dengan yang disukai pengguna. Dengan memanfaatkan kesamaan perilaku antar pengguna, CF dapat memberikan rekomendasi yang relevan dan personal.

III. METODOLOGI PENELITIAN

A. Dataset

Dataset yang digunakan pada penelitian ini adalah dataset Netflix Original Films & IMDB Scores yang diperoleh dari Kaggle. Kumpulan data ini mencakup semua film asli Netflix yang dirilis per 1 Juni 2021 dengan 584 baris data. Data dikumpulkan dari halaman Wikipedia dan dilengkapi dengan skor IMDB yang relevan. Skor IMDB ini dipilih oleh anggota komunitas dengan sebagian besar film memiliki lebih dari 1.000 ulasan. Data mencakup informasi sebagai berikut:

1. Judul film
2. Genre film
3. Tanggal tayang perdana asli
4. Durasi dalam hitungan menit
5. Skor IMDB (per 1 Juni 2021)
6. Bahasa yang tersedia saat ini (mulai 1 Juni 2021)

B. Rumus Matriks Kesamaan

Menghitung kesamaan antara pengguna berdasarkan rating film yang diberikan oleh pengguna lain. Kemudian menggunakan metode *cosine similarity* untuk mengukur kesamaan antara pengguna. Salah satu tahap penting dalam algoritma *item-based collaborative filtering* adalah menghitung kemiripan (*similarity*) antara item-item dan memilih item yang paling mirip. Dalam proses ini, digunakan persamaan untuk menghitung nilai kemiripan antara item tertentu i dan pengguna u . *Cosine similarity* digunakan sebagai metode untuk melakukan perhitungan tingkat kemiripan dari data. Berikut rumus yang digunakan oleh *cosine similarity*.

$$\text{sim}(i, u) = \cos(\vec{i}, \vec{u}) = \frac{\vec{i} \cdot \vec{u}}{\|\vec{i}\|_2 \|\vec{u}\|_2}$$

Keterangan:

\vec{i} = vektor i , yang akan dibandingkan kemiripannya

\vec{u} = vektor u , yang akan dibandingkan kemiripannya

$\|\vec{i}\|$ = panjang vektor item i

$\|\vec{u}\|$ = panjang vektor pengguna u

Setelah mengidentifikasi sekelompok item yang mirip dengan menggunakan rumus *cosine similarity*, langkah berikutnya adalah melakukan proses perhitungan prediksi yang memperkirakan nilai rating dari pengguna untuk item yang belum pernah diberi rating. Jika nilai prediksi tersebut tinggi, hal ini mengindikasikan bahwa target pengguna mungkin tertarik dengan item tersebut. Teknik yang digunakan untuk menghitung nilai prediksi ini adalah dengan menggunakan persamaan *weighted sum*.

$$P_{u,i} = \frac{\sum_{i \in I} (S_{i,u} * R_{u,i})}{\sum_{i \in I} (|S_{i,u}|)}$$

Keterangan:

$P_{u,i}$ = Nilai prediksi rating untuk pengguna u pada item i

$\sum i \in I$ = Semua himpunan item yang mirip dengan item I

$R_{u,i}$ = Rating pengguna u pada item i

$S_{u,i}$ = Nilai kemiripan antara item i dan pengguna u yang mirip

IV. HASIL DAN PEMBAHASAN

A. Pembersihan Data

Beberapa judul film masih memiliki spasi tambahan yang tidak perlu dan kesalahan ejaan. Berikut adalah perbaikan yang dilakukan pada beberapa judul film:

- a. Mengganti "American Factory: A Conversation with the Obamas " menjadi "American Factory: A Conversation with the Obamas".
- b. Mengganti "Bigflo & Oil: Hip Hop Frenzy" menjadi "Bigflo & Oli: Hip Hop Frenzy".

B. Modifikasi Kolom

Pada tahap ini dilakukan penggantian nama dan penamabahan kolom baru.

- a. Kolom yang semula bernama "IMDB Score" diubah menjadi "Score" untuk keperluan analisis lebih lanjut.
- b. Kolom "Genre + Language" yang merupakan gabungan dari kolom "Genre" dan "Language".
- c. Mengganti semua kata "romantic" pada kolom "Genre" dengan kata "romance".

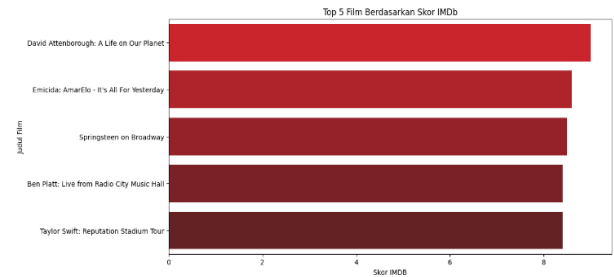
C. Visualisasi Top 5 Film Berdasarkan Skor IMDB

Pada Gambar 4.1 dan 4.2, ditampilkan lima film teratas berdasarkan skor IMDB

Gambar 4.1 Code untuk Visualisasi

```
top_movies = movies.nlargest(5, 'Score')
plt.figure(figsize=(12, 6))
sns.barplot(x='Score', y='Title', data=top_movies, palette=netflix_palette)
plt.title('Top 5 Film Berdasarkan Skor IMDB')
plt.xlabel('Skor IMDB')
plt.ylabel('Judul Film')
plt.show()
```

Gambar 4.2 Visualisasi Top 5 Film berdasarkan Score IMDB



D. Perhitungan Matriks Kesamaan

Pada penelitian ini, dilakukan perhitungan kesamaan untuk rekomendasi film menggunakan *cosine similarity* antara film berdasarkan kolom "Genre + Language" dan "Genre". Berikut adalah langkah-langkah yang dilakukan:

- a. Tokenisasi Kustom: Fungsi *custom_tokenizer* digunakan untuk memisahkan teks menjadi token-token terpisah. Tokenisasi membantu dalam memecah teks menjadi unit yang lebih kecil untuk pemrosesan lebih lanjut.
- b. Representasi Numerik: Kolom "Genre + Language" dan "Genre" dikonversi menjadi representasi numerik menggunakan *CountVectorizer*.

CountVectorizer menghitung frekuensi setiap token dalam teks dan mengubahnya menjadi vektor.

- c. Kombinasi Matriks: Matriks representasi numerik dari kolom "Genre + Language" dan "Genre" digabungkan menjadi satu matriks gabungan.
- d. Perhitungan Kesamaan: Matriks kesamaan dihitung menggunakan *cosine similarity*. *Cosine similarity* mengukur sudut antara dua vektor, di mana nilai 1 berarti vektor tersebut identik dan nilai 0 berarti vektor tersebut tidak memiliki kesamaan. Berikut adalah hasil sebagian dari perhitungan matriks kesamaan:

Gambar 4.3 Perhitungan Matriks Kesamaan

```

similarity_matrix

array([[1. , 0. , 0. , ..., 0. , 0.5, 0.5],
       [0. , 1. , 0. , ..., 0. , 0. , 0. ],
       [0. , 0. , 1. , ..., 0. , 0. , 0. ],
       ...,
       [0. , 0. , 0. , ..., 1. , 0. , 0. ],
       [0.5, 0. , 0. , ..., 0. , 1. , 0.5],
       [0.5, 0. , 0. , ..., 0. , 0.5, 1. ]])

```

E. Rekomendasi Film

Berikut adalah salah satu implementasi rekomendasi film berdasarkan kesamaan genre dan bahasa serta skor IMDb pada film “The Irishman” yang menghasilkan beberapa rekomendasi, diantaranya:

Tabel 4.1 Rekomendasi film “The Irishman”

No.	Judul Film	Skor	Kesamaan
1.	El Camino: A Breaking Bad Movie	7.3	1.0
2.	The Highwaymen	6.9	1.0
3.	Lost Girls	6.1	1.0
4.	Òlòtūré	5.5	1.0
5.	On My Skin	7.3	0.75

F. Hasil dan Diskusi

Hasil analisis menunjukkan bahwa film dengan genre dan bahasa yang serupa memiliki tingkat kesamaan yang tinggi. Hal ini terbukti dari rekomendasi yang dihasilkan untuk film “The Irishman”. Beberapa film seperti “The Highwaymen” dan “El Camino: A Breaking Bad Movie” menunjukkan kesamaan yang sempurna dengan film yang diminta, mengindikasikan bahwa metode *cosine similarity* efektif dalam mengidentifikasi film-film yang memiliki kesamaan tinggi berdasarkan genre dan bahasa. Selain itu, visualisasi lima film teratas berdasarkan skor IMDb memberikan wawasan tentang film-film terbaik di Netflix menurut penilaian IMDb. Analisis ini dapat membantu pengguna menemukan film-film dengan rating tinggi yang menarik.

V. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa metode *Collaborative Filtering* berbasis item dengan *cosine similarity* efektif untuk rekomendasi film di Netflix. Dataset yang mencakup semua film asli Netflix hingga 1 Juni 2021 dianalisis dengan teknik tokenisasi dan representasi numerik untuk genre dan bahasa. Hasilnya menunjukkan bahwa *cosine similarity* dapat mengidentifikasi film-film dengan kemiripan tinggi secara akurat, memberikan rekomendasi yang relevan seperti film “El Camino: A Breaking Bad Movie” dan “The Highwaymen” untuk “The Irishman”. Teknik ini, yang juga memanfaatkan prediksi berbasis *weighted sum*, membantu pengguna menemukan film yang sesuai dengan preferensi mereka, meningkatkan pengalaman menonton di Netflix.

REFERENCES

- [1] Survyana, Indah, “Mesin Rekomendasi Film Menggunakan Metode Kemiripan Genre Berbasis *Collaborative Filtering*”, Tesis, 2017.
- [2] Visher, Y., Bambang, S., dan Leopoldus, R., “Penerapan Metode *Item-Based Collaborative Filtering* Untuk Sistem Rekomendasi Data MovieLens”, Jurnal, p-ISSN:2302-4224, 2020.
- [3] Nugraha, D., Tito, W., dan Ratna, A., “Sistem Rekomendasi Film Menggunakan Metode *User Based Collaborative Filtering*”, Jurnal, Vol.8, No.5, Oktober 2021.
- [4] Hidayat, H. dan Agung, T., “Sistem Rekomendasi Film Menggunakan Metode Hybrid *Collaborative Filtering* Dan *Content-Based Filtering*”, Jurnal, Vol.9, No.3, Juni 2022.