Jurnal Ilmiah Teknik Informatika dan Komunikasi
Volume 5. Nomor. 3 November 2025
E-ISSN: 2827-7945 ; P-ISSN: 2827-8127, Hal 69-82
DOI: https://doi.org/10.55606/juitik.v5i3.1494
Available online at: https://journal.sinov.id/index.php/juitik

Penerapan *Content-Based Filtering* untuk Rekomendasi Film Netflix Berdasarkan Genre dan Klasifikasi Usia Penonton

Nayottama Nirpataka
Universitas Amikom Yogyakarta, Indonesia

Jl. Ring Road Utara, Sleman, Daerah Istimewa Yogyakarta
Korespondensi penulis: : nayottama12@students.amikom.ac.id

Abstract. *Recommendation systems have become an essential part of improving user experience, particularly in movie selection. With the ever-increasing number of movie titles, users often experience difficulty in determining which movies suit their interests and needs. Mismatches in movie selection, particularly in terms of age classification, can have negative impacts, particularly for children exposed to inappropriate content. Therefore, a recommendation system is needed that not only considers user preferences but also takes into account age restrictions. This study aims to build a movie recommendation system using a content-based filtering approach, considering two main aspects: genre and age classification. In its implementation, the Term Frequency-Inverse Document Frequency (TF-IDF) method is used as a word weighting technique for movie content information. This weighting is then used to calculate the level of similarity between movies using the Cosine Similarity method. This system is designed to be able to recommend relevant movies based on user preference input, such as the desired movie genre and the appropriate age classification. An evaluation was conducted to measure the extent to which the system is able to provide recommendations that match user preferences. The evaluation results show a similarity value of 1.00 for the genre aspect and 1.00 for the age classification aspect. This score indicates that the system successfully recommended highly relevant films that met the user's specified criteria. Therefore, the developed recommendation system effectively filters and suggests films that are not only content-appropriate but also safe for viewing based on the specified age restrictions.*

Keywords: *Content-Based Filtering, Cosine Similarity, Movie, Recommendation System, TF-IDF*

Abstrak. Sistem rekomendasi telah menjadi bagian penting dalam meningkatkan pengalaman pengguna, khususnya dalam pemilihan film. Dengan jumlah judul film yang terus bertambah, pengguna sering mengalami kesulitan dalam menentukan film mana yang sesuai dengan minat dan kebutuhan mereka. Ketidaksesuaian dalam pemilihan film, terutama dari segi klasifikasi usia, dapat berdampak negatif, terutama bagi anak-anak yang terpapar konten tidak sesuai. Oleh karena itu, diperlukan sistem rekomendasi yang tidak hanya mempertimbangkan preferensi pengguna tetapi juga memperhatikan faktor batasan usia. Penelitian ini bertujuan untuk membangun sistem rekomendasi film menggunakan pendekatan *content-based filtering* dengan mempertimbangkan dua aspek utama, yaitu genre dan klasifikasi usia. Dalam implementasinya, digunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) sebagai teknik pembobotan kata terhadap informasi konten film. Pembobotan ini kemudian digunakan untuk menghitung tingkat kemiripan antar film menggunakan metode *Cosine Similarity*. Sistem ini dirancang agar mampu merekomendasikan film yang relevan berdasarkan input preferensi pengguna, seperti genre film yang diminati serta klasifikasi usia yang sesuai. Evaluasi dilakukan untuk mengukur sejauh mana sistem mampu memberikan hasil rekomendasi yang sesuai dengan preferensi pengguna. Hasil evaluasi menunjukkan nilai kemiripan sebesar 1.00 untuk aspek genre dan 1.00 untuk aspek klasifikasi usia. Nilai tersebut menandakan bahwa sistem berhasil merekomendasikan film yang sangat relevan dan sesuai dengan kriteria yang telah ditentukan oleh pengguna. Dengan demikian, sistem rekomendasi yang dikembangkan ini efektif dalam menyaring dan menyarankan film yang tidak hanya sesuai secara konten, tetapi juga aman ditonton berdasarkan batasan usia yang ditentukan.

Kata kunci: Content Based Filtering, Cosine Similarity, Film, Sistem Rekomendasi, TF-IDF

1. LATAR BELAKANG

Industri perfilman di era digital seperti saat ini mengalami peningkatan penonton terutama pada platform streaming digital SVoD atau Streaming Video on Demand. Di Indonesia sendiri platform streaming berkembang sangat pesat ditandai dengan meningkatnya pelanggan layanan platform streaming digital. Menurut Bettina (2023) pelanggan layanan

Received: Juni 15, 2025; Revised: Juli 14, 2025; Accepted: Agustus 05, 2025; Online Available: Agustus 07, 2025

platform streaming digital pada 2016 sebanyak 200 ribu orang sementara di 2021 jumlahnya mencapai 11,5 juta pelanggan.

Netflix sebagai salah satu penyedia layanan streaming digital terbesar kedua memiliki lebih dari 10 ribu judul film, ini membuat penonton kesulitan dalam menentukan judul film yang sesuai dengan keinginan (Raihan, Najwatul, Selfi, Muhammad, and Adidtya 2025). Selain itu, pemilihan film yang tidak sesuai dengan umur akan menimbulkan resiko, misalnya memberikan tontonan film dengan rating dewasa kepada anak - anak akan menimbulkan dampak negatif dan membuat perilaku anak menjadi buruk (Ismatun, Isnatul, Luthfiyah, and Muhammad 2024).

Untuk menjawab tantangan tersebut, sistem rekomendasi menjadi salah satu solusi yang efektif untuk mengatasinya (Desvio and Eri 2025). Salah satu pendekatan yang sering digunakan adalah pendekatan Content Based Filtering, sistem dengan pendekatan ini dirancang untuk memberikan rekomendasi guna membantu pengguna dalam menentukan keputusan yang sesuai dengan preferensi pribadinya (Jamilatun, Vihi, and Nugroho 2024). Content Based Filtering akan merekomendasikan item sesuai dengan atribut film yang mencakup judul, genre, klasifikasi usia, pemain, negara asal, durasi, atau sutradara.

Sistem Rekomendasi dengan pendekatan Content Based Filtering sudah banyak diterapkan pada berbagai sistem karena kelebihannya. Pendekatan ini memiliki kelebihan yaitu tidak bergantung kepada pengguna atau preferensi dari user lain (Jamilatun, Vihi, and Nugroho 2024). Lalu, untuk algoritma yang akan diterapkan pada penelitian kali ini adalah TF-IDF dan Cosine Similarity. TF-IDF akan digunakan untuk melakukan pembobotan kata yang sering keluar, sedangkan cosine similarity akan digunakan untuk menghitung nilai cosinus pada sudut kata yang memiliki kesamaan (Jeremia, Lenny, Khoirudin, April, and Dede 2024).

Meskipun sistem rekomendasi dengan pendekatan Content Based Filtering sudah banyak diteliti seperti pada penelitian dengan judul rancang bangun sistem rekomendasi pemilihan drama korea dengan metode content-based filtering (Jamilatun, Vihi, and Nugroho 2024), sistem rekomendasi film pada platform streaming menggunakan metode content-based filtering (Jeremia, Lenny, Khoirudin, April, and Dede 2024), dan sistem rekomendasi film menggunakan metode content based filtering (Desvio end Eri 2025). Tidak banyak yang menggabungkan antara atribut genre dengan klasifikasi usia sebagai dasar untuk rekomendasinya. Padahal kedua atribut sangat penting untuk memastikan konten yang direkomendasikan tidak hanya sesuai dengan keinginan namun juga sesuai dengan batasan usia pengguna.Terutama pada anak - anak usia sekolah dasar yang memiliki resiko lebih besar pada perkembangannya (Ismatun, Isnatul, Luthfiyah, and Muhammad 2024). Oleh karena itu,

pendekatan dengan kedua atribut tersebut masih sangat diperlukan. Juga, diharapkan mampu memberikan rekomendasi yang tidak hanya relevan dengan preferensi pengguna, tetapi juga aman dan sesuai dengan batasan usia, sehingga bisa meningkatkan pengalaman menonton bagi semua kelompok usia.

2. KAJIAN TEORITIS

Sistem Rekomendasi

Sistem rekomendasi merupakan cabang dari bidang kecerdasan buatan (*Artificial Intelligence*) dimana sistem rekomendasi merupakan bagian dari sistem pengambilan informasi (*Information Retrieval*), karena dari preferensi pengguna sistem dapat merekomendasikan atau memprediksi item yang disukai dengan algoritma yang dimiliki sesuai dengan pendekatannya. Dan item yang dimaskud pada penelitian kali ini adalah judul film, dimana preferensi penggunaan genre dan klasifikasi usia (Raihan, Najwatul, Selfi, Muhammad, and Adidtya 2025).

Pada garis besarnya sistem rekomendasi memiliki 2 pendekatan utama yaitu content based filtering dan collaborative filtering. Pada pendekatan content based filtering sistem akan merekomendasikan item yang atributnya sama atau mirip dengan atribut yang sesuai dengan preferensi pengguna, dengan mencocokan atribut pada masing - masing kontennya. Sementara itu, collaborative filtering akan bergantung pada aktivitas dari penggunanya ataupun aktivitas pengguna lain yang mirip, dengan berorientasi pada item tersebut (Edbert, Seng, and Yustinus 2016).

Content Based Filltering

Sistem rekomendasi dengan pendekatan content based filtering memprediksi dan merekomendasikan item dengan memanfaatkan berbagai sumber informasi. Juga, dengan memanfaatkan konsep filtering informasi, sistem akan melakukan prediksi dan rekomendasi yang relevan melalui pemberian peringkat pada item yang telah dimiliki pengguna. Berdasarkan ukuran kesamaan antar item, item dengan ukuran paling mirip akan direkomendasikan kepada pengguna (Putri 2019). Metode yang sering digunakan pada rekomendasi dengan pendekatan content based filtering meliputi TF-IDF, Klasifikasi Bayes, analisis klaster, jaringan saraf tiruan, dan pohon keputusan (Nailatul and Anief 2021).

Metode

Pada penelitian kali ini akan menggunakan TF-IDF untuk pembobotan katanya, tujuannya untuk memberikan nilai pentingnya sebuah term pada suatu dokumen terhadap kumpulan dokumen lainnya. Semakin suatu term jarang muncul pada dokumen, artinya

semakin besar nilai IDF nya yang menunjukkan bahwa term semakin informatif. Bobot dari nilai TF-IDF didapatkan dari perkalian antara nilai tf (Term Frequency) term pada dokumen, dengan nilai IDF (Inverse Document Frequency) term tersebut. Lalu, nilai akan dinormalisasi agar berada di dalam rentang [0,1]. (Muhammad, Putra, and Agus 2021).

Setelah dilakukannya pembobotan dengan TF-IDF dilakukan perhitungan kemiripan dengan cosine similarity. Cosine similarity digunakan untuk menghitung kemiripan dari 2 dokumen, yang nantinya akan digunakan sebagai peringkat pada sistem rekomendasi. Nilai dari similaritas kosinus akan menghasilkan antara 1 dan 0, semakin besar atau mendekati 1 maka peringkatnya akan lebih tinggi (Nailatul and Anief 2021).

Atribut

Genre sebagai salah satu atribut film yang akan digunakan dalam mencari rekomendasi pada penelitian ini, dapat diartikan sebagai jenis dari sekelompok film dengan karakter yang sama atau mirip. Genre memiliki fungsi untuk membantu mengelompokkan film untuk penonton agar mengetahui gambaran umum alur ceritanya (Muhammad and Yolanda 2019).

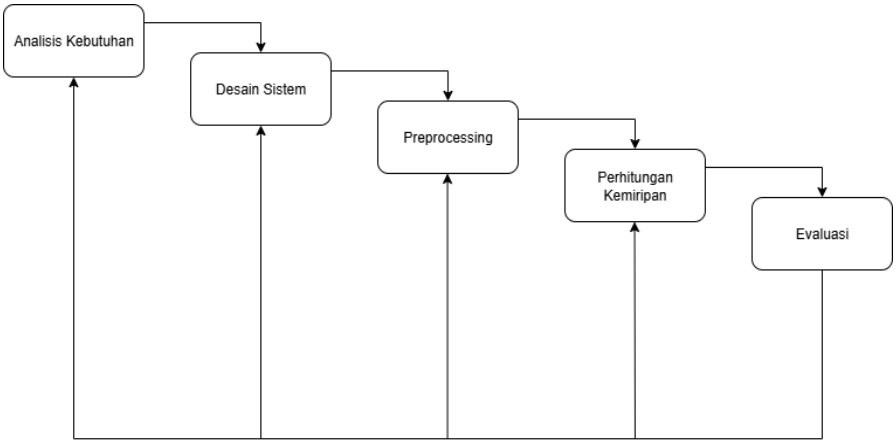
Selain genre, klasifikasi usia pada atribut film perlu diperhatikan khususnya untuk tontonan anak - anak. Klasifikasi usia pada film merupakan penyaringan konten berdasarkan rentang usia yang bisa dianggap cukup untuk suatu film. Pentingnya memberikan tontonan yang sesuai usia agar tidak terjadi hal negatif, seperti pada anak - anak karena anak - anak suka menirukan berbagai adegan yang ada di film (Rusnawati 2021).

Netflix

Netflix merupakan salah satu platform yang menyediakan layanan streaming film secara online yang bisa di askes kapanpun dan dimanapun seperti smartphone, televisi, pc, laptop, dan tablet. Penonton yang ingin menggunakan layanan netflix harus mendaftar sebagai member untuk mengakses acara TV dan filmnya, dimana penonton dapat melihat tayangan tanpa iklan dengan perangkat yang terhubung internet (Bernadus, Marcell, Ali, and Felix 2021).

3. METODE PENELITIAN

Pada penelitian ini akan digunakan metode waterfall, dikarenakan setiap tahapan yang ada pada metode ini harus diselesaikan secara berurutan. Pada setiap tahap yang akan dilakukan harus menunggu tahap sebelumnya selesai terlebih dahulu. Sehingga, waterfall dirasa cukup tepat penerapannya, selain itu juga metode ini sangat umum digunakan pada pengembangan software (Ryky, Bagus, and Mufti 2023). Gambaran alur metode waterfall dapat dilihat pada gambar 1 berikut.



Gambar 1. Metode Waterfall

Pada gambar 1 penjelasan setiap tahapnya :

- Analisis kebutuhan

  Pada tahap ini dilakukan identifikasi kebutuhan sistem rekomendasi , mulai dari dataset hingga atribut yang akan digunakan.

- Desain Sistem

  Berdasarkan hasil analisis kebutuhan, pada tahap ini akan dilakukan desain sistem mulai dari arsitektur sistem, dan pemodelan algoritma sistem.

- Preprocessing

  Tahap ini akan menghasilkan dataset yang sudah bersih dan siap digunakan, karena sebelumnya sudah dilakukan beberapa tahapan pra pemrosesan data.

- Perhitungan kemiripan

  Setelah dilakukan pembersihan data pada tahap sebelumnya, data akan dihitung kemiripannya dengan menggunakan TF-IDF dan Cosine Similarity.

- Evaluasi

  Pada tahapan terakhir, sistem akan dievaluasi untuk melihat rekomendasi yang dihasilkan sistem apakah sudah sesuai dan relevan dengan preferensi dari pengguna.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini merupakan sistem rekomendasi berbasis konten (Content Based Filtering) dengan metode TF-IDF dan Cosine Similarity. Sistem akan menampilkan hasil rekomendasi sesuai dengan genre dan klasifikasi usia yang diinputkan pengguna, berikut ini merupakan hasil dan pembahasan dari setiap tahapan dari metode pengembangan yang sudah dirancang.

Analisis Kebutuhan

Hasil dari identifikasi kebutuhan yang sudah dilakukan, akan digunakan dataset dari sumber terbuka dengan judul "Recflix: Sistem Rekomendasi Netflix" dimana dataset didapatkan dari situs web kaggle yang memuat 7787 data dengan 12 kolom atribut. Dari 12 atribut yang akan digunakan pada penelitian ini 7 atribut yaitu type, tittle, listed_in, rating, release_year, country, dan duration.

Jumlah judul di dataset: 7787

	show_id	type	title	director	cast	country	date_added	release_year	rating	duration	listed_in	description
0	s1	TV Show	3%	NaN	João Miguel, Bianca Comparato, Michel Gomes, R...	Brazil	August 14, 2020	2020	TV-MA	4 Seasons	International TV Shows, TV Dramas, TV Sci-Fi &...	In a future where the elite inhabit an island ...
1	s2	Movie	7:19	Jorge Michel Grau	Demián Bichir, Héctor Bonilla, Oscar Serrano, ...	Mexico	December 23, 2016	2016	TV-MA	93 min	Dramas, International Movies	After a devastating earthquake hits Mexico Cit...
2	s3	Movie	23:59	Gilbert Chan	Tedd Chan, Stella Chung, Henley Hii, Lawrence ...	Singapore	December 20, 2018	2011	R	78 min	Horror Movies, International Movies	When an army recruit is found dead, his fellow...
3	s4	Movie	9	Shane Acker	Elijah Wood, John C. Reilly, Jennifer Connelly...	United States	November 16, 2017	2009	PG-13	80 min	Action & Adventure, Independent Movies, Sci-Fi...	In a postapocalyptic world, rag-doll robots hi...
4	s5	Movie	21	Robert Luketic	Jim Sturgess, Kevin Spacey, Kate Bosworth, Aar...	United States	January 1, 2020	2008	PG-13	123 min	Dramas	A brilliant group of students become card-coun...

Gambar 2. Contoh Isi Dataset

Gambar 2 menampilkan contoh data yang ada di dalam dataset, dan jumlah data yang ada pada datasetnya yaitu sebanyak 7787 judul film dengan 12 atribut.

Desain Sistem

Pada tahap desain sistem akan ditentukan arsitektur sistem dengan 5 bagian, yang pertama input. Pengguna sistem akan melakukan input untuk memilih jenis tontonan, yang berupa pilihan antara film atau TV show. Setelah itu pengguna baru bisa melakukan input untuk genre dan klasifikasi usianya. Pada input klasifikasi usia, pengguna cukup melakukan input umur berupa angka. Untuk bagian kedua yaitu dataset dengan atribut yang digunakan yaitu type, title, genre, listed_in, duration, country, release_year. Lalu, bagian ketiga pre-processing data, akan dilakukan proses pembersihan data kosong, pembersihan duplikasi data, penggabungan atribut genre dan klasifikasi usia, pemisahan tipe film dan perhitungan TF-IDF. Pada bagian keempat merupakan perhitungan kemiripan antar film berdasarkan perhitungan TF-IDF sebelumnya. Dan pada bagian kelima yaitu output, sistem rekomendasi akan menampilkan daftar film sesuai dengan preferensi pengguna.

Lalu bagian pemodelan algoritma sistem rekomendasi akan dilakukan dengan langkah - langkah berikut :

- Pre-processing data

  Akan dilakukan pembersihan data kosong dan duplikat, lalu dilakukan penggabungan 2 kolom genre dan klasifikasi usia menjadi satu fitur untuk dilakukan vektorisasi, terakhir dilakukan pemisahan jenis atau tipe tontonan menjadi 2 yaitu movie

dan tv show sesuai yang ada di dataset. Gunanya untuk memastikan hasil rekomendasi yang sesuai fokus tipe atau jenisnya, karena tipe movie memiliki karakteristik berbeda dengan tv show.

- Vektorisasi TF-IDF

  Dengan menggunakan TF-IDF akan diubah dari teks menjadi angka numerik. Proses ini dilakukan untuk perhitungan pembobotan setiap kata yang ada di dalam fitur penggabungan genre dan klasifikasi usia.

- Perhitungan Cosine Similarity

  Untuk menghitung kemiripan antar film akan digunakan cosine similarity. Semakin tinggi nilai cosine yang dihasilkan maka semakin mirip dua film tersebut.

- Pemfilteran hasil rekomendasi

  Setelah dilakukan perhitungan, akan dilakukan pemfilteran untuk mengambil film yang sesuai dengan preferensi pengguna. Sehingga hanya film yang sesuai genre dan klasifikasi usia dari input pengguna saja yang ditampilkan.

- Output

  Akan menampilkan 10 rekomendasi film teratas berdasarkan dengan skor perhitungan similarity.

Preprocessing

Preprocessing dilakukan untuk membersihkan dan mempersiapkan data film agar siap digunakan, beberapa tahapan preprocessing akan dijelaskan dibawah.

```
print("Jumlah data sebelum hapus duplikat:", len(df))

df = df.drop_duplicates(subset=['title', 'release_year'], keep='first')

print("Jumlah data setelah hapus duplikat:", len(df))
```

```
Jumlah data sebelum hapus duplikat: 7787
Jumlah data setelah hapus duplikat: 7787
```

Gambar 3. Penghapusan data duplikat

Pada gambar 3 menampilkan kode untuk menghapus data duplikasi berdasarkan title dan release_year, gunanya untuk memastikan bahwa setiap judul film per tahun rilis, sehingga tidak akan menimbulkan rekomendasi yang bias. Dari total data sebelum dilakukan pembersihan dan setelah pembersihan tetap sama 7787, dengan begitu proses drop_duplicates() tidak mengubah jumlah data karena setiap judul per tahun rilisnya sudah bersih dari duplikasi.

```
print("Jumlah data sebelum drop NA:", len(df))

df_reco = df[['type', 'title', 'listed_in', 'rating', 'release_year', 'country', 'duration']].dropna()

print("Jumlah data setelah drop NA:", len(df_reco))
```

```
Jumlah data sebelum drop NA: 7787
Jumlah data setelah drop NA: 7274
```

Gambar 4. Penghapusan data kosong

Pada gambar 4 diatas menampilkan kode untuk melakukan pembersihan dataset dari nilai kosong agar hasil rekomendasi akurat. Pada kolom yang akan dipakai dilakukan proses seleksi dengan menggunakan dropna(). Dari total awal jumlah data 7787 menjadi 7274 data. Artinya terdapat 1020 baris yang memiliki informasi tidak lengkap.

```
print("Sebelum kolom 'features' ditambahkan:")
print(df_reco[['title', 'listed_in', 'rating']].head())

df_reco['features'] = df_reco['listed_in'] + ' ' + df_reco['rating']
df_reco['features'] = df_reco['features'].str.replace(',', ' ', regex=False).str.lower()

print("\nSetelah kolom 'features' ditambahkan:")
print(df_reco[['title', 'features']].head())
```

```
Sebelum kolom 'features' ditambahkan:
   title                                        listed_in rating
0     3%  International TV Shows, TV Dramas, TV Sci-Fi &...  TV-MA
1   7:19                  Dramas, International Movies  TV-MA
2  23:59               Horror Movies, International Movies      R
3      9  Action & Adventure, Independent Movies, Sci-Fi...  PG-13
4     21                                       Dramas  PG-13

Setelah kolom 'features' ditambahkan:
   title                                        features
0     3%  international tv shows  tv dramas  tv sci-fi &...
1   7:19               dramas  international movies tv-ma
2  23:59            horror movies  international movies r
3      9  action & adventure  independent movies  sci-fi...
4     21                                  dramas pg-13
```

Gambar 5. Penggabungan Atribut

Berdasarkan gambar 5 diatas, proses yang dilakukan adalah penggabungan dari atribut genre / kategori dan klasifikasi usia / rating untuk menjadikan satu informasi dan ditempatkan ke dalam kolom features. Dimana isi kolom feature merupakan data dari kategori dan rating dimana koma akan diganti menjadi spasi agak tidak dibaca sebagai token berbeda, serta semua teks akan diubah ke huruf kecil untuk dilakukan pengolahan data selanjutnya.

```
type_input = "Movie"
filtered_df = df_reco[df_reco['type'] == type_input]

print(f"Jumlah data Movie: {len(filtered_df)}")
filtered_df[['title', 'features']].head()
```

```
Jumlah data Movie: 5143
```

	title	features
1	7:19	dramas international movies tv-ma
2	23:59	horror movies international movies r
3	9	action & adventure independent movies sci-fi...
4	21	dramas pg-13
6	122	horror movies international movies tv-ma

Gambar 6. Pemisahan Movie dan TV Show

Dari gambar 6 diatas, dilakukan penyaringan untuk memisah antara tipe move dan tv show. Sebagai contoh akan digunakan type movie di mana hasilnya menunjukkan jumlah data 5143 yang memiliki type movie. Dimana selanjutnya akan dilakukan perhitungan TF-IDF nya.

```
[27] vectorizer = TfidfVectorizer(max_features=500)
     tfidf_matrix = vectorizer.fit_transform(filtered_df['features'])

     print("Contoh fitur (vocab):")
     print(vectorizer.get_feature_names_out()[:30])  # ambil 30 fitur pertama
```

```
Contoh fitur (vocab):
['13' '14' '17' 'action' 'adventure' 'anime' 'children' 'classic'
 'comedies' 'comedy' 'cult' 'documentaries' 'dramas' 'faith' 'family'
 'fantasy' 'features' 'fi' 'fv' 'horror' 'independent' 'international'
 'lgbtq' 'ma' 'movies' 'music' 'musicals' 'nc' 'nr' 'pg']
```

Gambar 7. Perhitungan TF-IDF

Gambar 7 menampilkan perhitungan atau pembobotan teks menjadi angka numerik dengan TF-IDF Vectorizer. Dari seluruh data movie akan diambil 500 kata (fitur) paling relevan. Metode TF-IDF ini akan menghitung bobot pentingnya setiap kata dalam masing-masing film, sehingga sistem dapat mengenali ciri khas genre ataupun rating usianya.

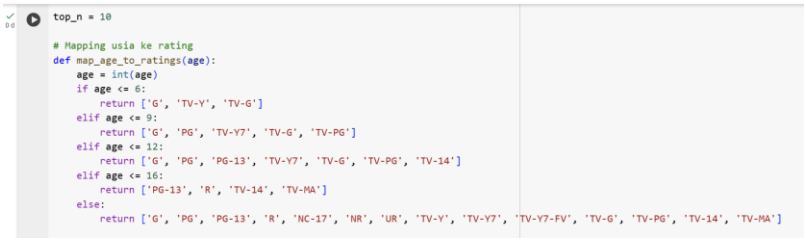
Perhitungan Kemiripan

Sistem akan melakukan perhitungan kemiripan, dengan menggunakan cosine similarity. Dimana hasil input preferensi pengguna akan dihitung dengan data yang berada di dalam dataset.

```
# Contoh input dari user
type_input = "Movie"
genre_input = "Anime Features"
age_input = 8
```

Gambar 8. Contoh Input dari Pengguna

Gambar 8 merupakan contoh input data untuk rekomendasi yang sudah dilakukan oleh pengguna. Pengguna melakukan untuk pilih type tontonan yaitu “Movie”, lalu genre “Anime Features”, dan usia “8”.

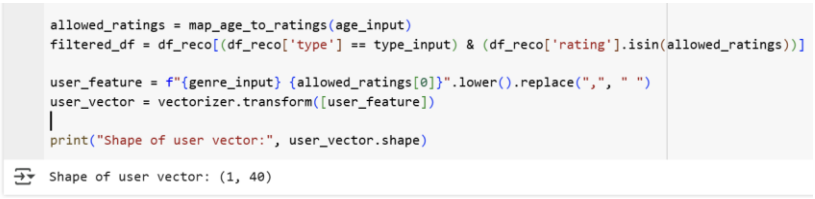


```
top_n = 10

# Mapping usia ke rating
def map_age_to_ratings(age):
    age = int(age)
    if age <= 6:
        return ['G', 'TV-Y', 'TV-G']
    elif age <= 9:
        return ['G', 'PG', 'TV-Y7', 'TV-G', 'TV-PG']
    elif age <= 12:
        return ['G', 'PG', 'PG-13', 'TV-Y7', 'TV-G', 'TV-PG', 'TV-14']
    elif age <= 16:
        return ['PG-13', 'R', 'TV-14', 'TV-MA']
    else:
        return ['G', 'PG', 'PG-13', 'R', 'NC-17', 'NR', 'UR', 'TV-Y', 'TV-Y7', 'TV-Y7-FV', 'TV-G', 'TV-PG', 'TV-14', 'TV-MA']
```

Gambar 9. Klasifikasi Usia

Gambar 9 merupakan mapping usia yang sudah disesuaikan dengan kategori rating usia yang digunakan di dalam dataset netflix. Dan untuk memastikan konten yang direkomendasikan sesuai usia digunakan klasifikasi rating secara internasional yang mengacu pada pengaduan dari Motion Picture Association dan TV Parental Guidelines.



```
allowed_ratings = map_age_to_ratings(age_input)
filtered_df = df_reco[(df_reco['type'] == type_input) & (df_reco['rating'].isin(allowed_ratings))]

user_feature = f"{genre_input} {allowed_ratings[0]}".lower().replace(",", " ")
user_vector = vectorizer.transform([user_feature])
print("Shape of user vector:", user_vector.shape)

Shape of user vector: (1, 40)
```

Gambar 10. Pembentukan Fitur Pengguna

Gambar 10 dilakukan proses filter data berdasarkan usia dan tipe yang diinputkan oleh pengguna. Lalu genre yang diinputkan pengguna akan digabung dengan rating yang sesuai mapping dan dijadikan satu dalam dokumen fitur. Lalu dokumen fitur akan dilakukan pembobotan dengan TF-IDF.



```
from IPython.display import display

if not filtered_df.empty:
    filtered_df = filtered_df.reset_index(drop=True)

    tfidf_matrix = vectorizer.fit_transform(filtered_df['features'])

    user_feature = f"{genre_input} {' '.join(allowed_ratings)}".lower().replace(",", " ")
    user_vector = vectorizer.transform([user_feature])

    similarity_scores = cosine_similarity(user_vector, tfidf_matrix).flatten()

    top_n = 10
    available_n = min(top_n, len(similarity_scores))
    top_indices = similarity_scores.argsort()[-available_n:][::-1]

    recommended_df = filtered_df.iloc[top_indices].copy()
    recommended_df['similarity_score'] = similarity_scores[top_indices]

    display_cols = ['title', 'listed_in', 'rating', 'release_year', 'country', 'duration', 'similarity_score']
    recommended_df = recommended_df[display_cols]

    print("\n Top Rekomendasi Berdasarkan Genre & Usia:")
    display(recommended_df)
```

Gambar 11. Perhitungan Kemiripan

Gambar 11 merupakan proses perhitungan kesamaan antara preferensi pengguna dengan datanya. Dari pembentukan fitur pengguna, hasilnya akan dihitung kemiripannya dengan seluruh data. Lalu akan dipilih 10 perhitungan yang paling tinggi atau paling mendekati 1.

Top Rekomendasi Berdasarkan Genre & Usia:

	title	listed_in	rating	release_year	country	duration	similarity_score
160	Children of the Sea	Anime Features	TV-PG	2019	Japan	112 min	0.855050
933	Yo-Kai Watch: The Movie	Anime Features, Children & Family Movies	TV-Y7	2016	Japan, United States	96 min	0.844308
634	Pokémon: Mewtwo Strikes Back - Evolution	Anime Features, Children & Family Movies	TV-Y7	2019	Japan	98 min	0.844308
632	Pokémon the Movie: I Choose You!	Anime Features, Children & Family Movies	TV-Y7	2017	Japan	97 min	0.844308
633	Pokémon the Movie: Power of Us	Anime Features, Children & Family Movies	TV-Y7	2018	Japan	98 min	0.844308
261	Fireworks	Anime Features, Romantic Movies	TV-PG	2017	Japan	91 min	0.759473
39	A Whisker Away	Anime Features, Children & Family Movies, Inte...	TV-PG	2020	Japan	104 min	0.725954
505	Modest Heroes: Ponoc Short Films Theatre	Anime Features, Children & Family Movies	PG	2018	Japan	54 min	0.708940
589	Okko's Inn	Anime Features, Children & Family Movies	PG	2018	Japan	96 min	0.708940
479	Mary and the Witch's Flower	Anime Features, Children & Family Movies	PG	2017	Japan	103 min	0.708940

Gambar 12. Hasil Rekomendasi

Gambar 12 merupakan hasil rekomendasi dari sistem sesuai dengan preferensi yang diinputkan oleh pengguna. Pengguna melakukan input type “Movie”, genre “ Anime Features”, dan uisa “8”. Output yang ditampilkan sudah dipisahkan untuk movie dan TV Show, jadi pada hasil ini menampilkan film dengan type Movie saja. Pada tabel bagian genre dapat dilihat bahwa hasilnya menunjukkan semua film memiliki genre yang sama dengan yang diinputkan pengguna. Untuk rating usia ditampilkan TV-Y7, TV-PG, dan PG, dimana TV-Y7 cocok untuk anak 7 tahun keatas sehingga cocok untuk usia 8 tahun, sedangkan TV-PG dan PG disarankan dengan bimbingan orang tua namun ini juga cukup cocok untuk usia 8 tahun. Untuk nilai similarity tertingginya 0.855050, menandakan bahwa hasil rekomendasi memiliki kemiripan tinggi dengan preferensi pengguna.

Evaluasi

Evaluasi sistem dilakukan untuk mengetahui apakah genre dan usia yang direkomendasikan sudah sesuai dengan preferensi dari pengguna.

```
[43] def evaluate_genre_precision_at_k(recommended_df, user_genre, k=10):
         user_keywords = user_genre.lower().strip().split()

         def genre_match(listed_in):
             listed = listed_in.lower()
             return any(keyword in listed for keyword in user_keywords)

         actual_k = min(k, len(recommended_df))
         if actual_k == 0:
             return 0.0

         match_count = recommended_df.head(actual_k)['listed_in'].apply(genre_match).sum()
         precision = match_count / actual_k
         return precision

     genre_precision = evaluate_genre_precision_at_k(recommended_df, genre_input, top_n)
     print(f"\n Precision@{top_n} Genre: {genre_precision:.2f}")

 Precision@10 Genre: 1.00
```

Gambar 13. Evaluasi Genre

Pada gambar 13 menunjukkan hasil evaluasi dengan menggunakan metrik Precision@10 dimana hasilnya 1.00 atau 100%, yang berarti menunjukkan kesesuaian antara hasil 10 teratas

dengan input pengguna. Evaluasi dilakukan dengan mencocokan kolom listed_in dengan input genre dari pengguna.

```python
# Evaluasi precision berdasarkan kesesuaian rating usia
def evaluate_age_precision(recommended_df, allowed_ratings):
    if len(recommended_df) == 0:
        return 0.0

    match_count = recommended_df['rating'].isin(allowed_ratings).sum()
    precision = match_count / len(recommended_df)
    return precision

age_precision = evaluate_age_precision(recommended_df, allowed_ratings)

print(f" Precision Usia: {age_precision:.2f}")
```

```
 Precision Usia: 1.00
```

Gambar 14. Evaluasi Rating Usia

Pada gambar 14 menunjukkan hasil evaluasi kesesuain rating usia dengan menggunakan Precision Usia dimana hasilnya 1.00, artinya semua film yang ditampilkan pada hasil rekomendasi sudah sesuai batasan rating usia yang diperbolehkan, ditonton oleh usia yang sudah diinputkan pengguna. Evaluasi dihitung dengan membandingkan rating hasil rekomendasi dengan daftar rating yang diizinkan.

5. KESIMPULAN DAN SARAN

Dari hasil evaluasi yang sudah dilakukan pada sistem hasilnya menunjukkan nilai precision yang tinggi dan baik pada genre maupun klasifikasi usia. Hal ini menunjukan bahwa sistem berhasil merekimendasikan film berdasarkan genre dan klasifikasi usia sesuai dengan preferensi dari pengguna dan batasan usia yang ditetapkan. Evaluasi dilakukan dengan metrik precision, dengan hasil yang menunjukkan bahwa pendekatan yang digunakan pada penelitian ini relevan dan efektif sesuai dengan kebutuhan. Meskipun hasil dari evaluasi sistem sudah sangat baik, pengujian lebih lanjut disarankan dengan beragam skenario, untuk mengetahui apakah sistem memang benar - benar bekerja dengan baik. Selain itu, pengembangan fitur evaluasi seperti recall atau user satisfaction score juga perlu diperitmbangakn untuk mengetahui keseluruhan gambaran sistem rekomendasi.

DAFTAR REFERENSI

Ardiansyah, R., Bianto, M. A., & Saputra, B. D. (2023). Sistem rekomendasi buku perpustakaan sekolah menggunakan metode content-based filtering. *Jurnal CoSciTech (Computer Science and Information Technology)*, *4*(2), 510–518. https://doi.org/10.37859/coscitech.v4i2.5131

Avinger, C. L. (2022). TV Parental Guidelines system. *EBSCO*. Retrieved July 29, 2025, from https://www.ebsco.com/research-starters/communication-and-mass-media/tv-parental-guidelines-system

Azizah, N., & Rozi, A. F. (2024). Sistem rekomendasi produk Somethinc menggunakan metode content-based filtering. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *6*(3), 461–468. https://doi.org/10.47233/jteksis.v6i3.1411

Azri Saputra, J. M., Huizen, L. M., & Arianto, D. B. (2024). Sistem rekomendasi film pada platform streaming menggunakan metode content-based filtering. *Jurnal Transformatika*, *22*(1), 10–21. https://doi.org/10.26623/transformatika.v22i1.7041

Hadi, G. N. (2019). Sistem rekomendasi film menggunakan metode content-based filtering. *Skripsi*. http://e-journal.uajy.ac.id/7244/4/3TF03686.pdf

Insan, R., Siagian, P., Khoiriah, N., Priscilia, S. A., Perdana, A., & Komputer, I. (2025). Penerapan machine learning untuk rekomendasi film berdasarkan preferensi pengguna. *JATI (Jurnal Teknik Informatika)*, *9*(4), 5658–5662. https://doi.org/10.36040/jati.v9i4.13884

Ismatun Nisak, I., Mukarromah, I., Aditama, L. M., & Zulfahmi, M. N. (2024). Pentingnya filterisasi konten dewasa pada perkembangan sosial emosional anak sekolah dasar. *Jurnal Bintang Pendidikan Indonesia*, *3*(1), 199–209. https://doi.org/10.55606/jubpi.v3i1.3580

Nastiti, P. (2019). Penerapan metode content based filtering dalam implementasi sistem rekomendasi tanaman pangan. *Teknika*, *8*(1), 1–10. https://doi.org/10.34148/teknika.v8i1.139

Rizky, M. Y., & Stellarosa, Y. (2019). Preferensi penonton terhadap film Indonesia. *Communicare: Journal of Communication Studies*, *4*(1), 15. https://doi.org/10.37535/101004120172

Rochmad Wahono, A., Saputra, B. A., & Rahman, F. F. (2024). Sistem rekomendasi film menggunakan metode content-based filtering dan algoritma K-Nearest Neighbors (KNN). *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis*, 1–6. https://doi.org/10.47701/senatib.v4i1.3994

Rusnawati, R. (2021). Efektivitas sensor mandiri pada orang tua terhadap tontonan anak usia 2–6 tahun. *JURNAL HURRIAH: Jurnal Evaluasi Pendidikan dan Penelitian*, *2*(4), 108–113. https://doi.org/10.56806/jh.v2i4.37

Safitri, J., Atina, V., & Sudibyo, N. A. (2024). Rancang bangun sistem rekomendasi pemilihan drama Korea dengan metode content-based filtering. *Infotech*, *5*(2), 175–189. https://doi.org/10.37373/infotech.v5i2.1235

Salim, E., Pragantha, J., & Manatap, D. L. (2021). Perancangan sistem rekomendasi film menggunakan metode content-based filtering. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, *5*(6), 2188–2199. https://lintar.untar.ac.id/repository/penelitian/buktipenelitian_10390001_7A281222103549.pdf

Shafina, G. (2023, July 19). Jumlah pelanggan layanan streaming Indonesia kian meningkat. *GoodStats Data*. Retrieved July 29, 2025, from https://data.goodstats.id/statistic/jumlah-pelanggan-layanan-streaming-indonesia-kian-meningkat-s2q42

Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis data mining data Netflix menggunakan aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems*, *4*(1), 13–21. https://doi.org/10.30813/jbase.v4i1.2729

Sumarlin, E. W., Hansun, S., & Wiratama, Y. W. (2016). Rancang bangun aplikasi rekomendasi film dengan menggunakan algoritma Simple Additive Weighting. *Jurnal Informatika, 10*(2), 1244–1250. https://doi.org/10.26555/jifo.v10i2.a5066