



Sistem Rekomendasi Film pada *Platform Streaming* Menggunakan Metode *Content-Based Filtering*

Jeremia Maheswara Azri Saputra^{1*}, Lenny Margaretta Huizen², Khoirudin³, April Firmandaru⁴, Dede Brahma Arianto⁵

¹Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, RT.7/RW.7, Tlogosari Kulon, Kec. Pedurungan, Kota Semarang, Jawa Tengah 50196, e-mail: jmahesazri@gmail.com

²Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, RT.7/RW.7, Tlogosari Kulon, Kec. Pedurungan, Kota Semarang, Jawa Tengah 50196, e-mail: lenny@usm.ac.id

³Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, RT.7/RW.7, Tlogosari Kulon, Kec. Pedurungan, Kota Semarang, Jawa Tengah 50196, e-mail: khoirudin@usm.ac.id

⁴Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
Jl. Soekarno Hatta, RT.7/RW.7, Tlogosari Kulon, Kec. Pedurungan, Kota Semarang, Jawa Tengah 50196, e-mail: firman@usm.ac.id

⁵Universitas Islam Indonesia
Jl. Kaliurang No.Km. 14,5, Krawitan, Umbulmartani, Ngemplak, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55584, e-mail: dede.brahma2@gmail.com

ARTICLE INFO

History of the article :

Received 9 Juni 2023

Received in revised form 10 Juni 2024

Accepted 12 Juli 2024

Available online 8 Agustus 2024

Keywords:

Film Recommendation System; Streaming Platform; Content-Based Filtering; TF-IDF; Cosine Similarity

* Correspondence:

Telepon:

-

E-mail:

jmahesazri@gmail.com

ABSTRACT

The development of digital technology has affected many aspects of life, including the film industry. Nowadays, watching films can be done through streaming platforms. However, with a lot of films available on the streaming platform, it makes the audiences take a long time in choosing films. In that case, the aim of this research is to implement the Content-Based Filtering method to find similarities of films by calculating the word weight of a movie through the description of a film using the TF-IDF and Cosine Similarity methods, thus creating a Film Recommendation System that can help people in finding films similar to their favorite films, the Film Recommendation System in this research is divided into movie type recommendation system and show type recommendation system. The result of this research is the creation of a Film Recommendation System on the streaming platform with accuracy value is 0.7022 for movie type and 0.7973 for show type.

1. INTRODUCTION

Perkembangan teknologi digital telah mempengaruhi banyak aspek kehidupan, salah satunya di bidang industri film. Industri perfilman merupakan salah satu sektor industri budaya yang memiliki pasar yang luas secara global [1]. Demi memenuhi pangsa pasar industri film, dengan berkembangnya teknologi digital di industri film, saat ini menonton film dapat dilakukan dimana

saja dan kapan saja melalui platform *streaming*. Platform *streaming* adalah sebuah platform digital berbentuk *website* dan aplikasi yang berisikan banyak film yang dapat langsung ditonton oleh pengguna secara *online*. Akan tetapi, banyaknya pilihan film yang tersedia juga membuat para penonton membutuhkan lebih banyak perhatian dalam memilih film yang akan ditonton, akibatnya banyak pengguna yang menghabiskan lebih banyak waktu dalam menentukan pilihan film yang akan ditonton. Penyebab masalah para penonton dalam menentukan pilihan film yang akan ditonton adalah karena banyaknya pilihan film dari berbagai genre yang dapat ditonton [2], para penonton membutuhkan waktu lebih banyak dalam menyesuaikan selera mereka kedalam pilihan film yang akan mereka tonton, ataupun mencari film yang mirip dengan film kesukaan penonton.

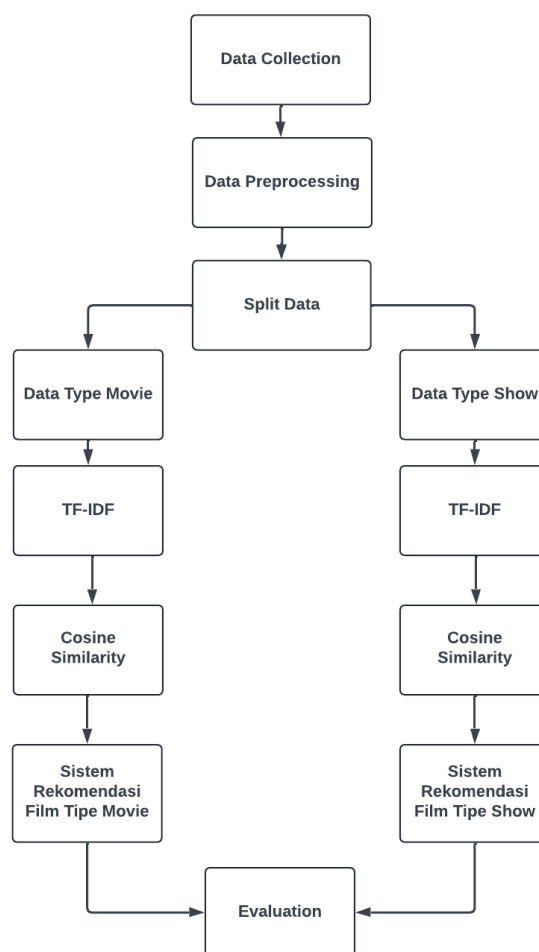
Berdasarkan situasi di atas, solusi untuk mengatasi permasalahan pengguna platform *streaming* yaitu dengan membuat suatu sistem rekomendasi film pada platform *streaming*, sehingga nantinya para penonton tidak menghabiskan waktu yang lama dalam mencari film yang akan ditonton melalui platform *streaming*. Sistem rekomendasi film dapat dibuat dengan beberapa metode, seperti metode *content-based filtering*, metode *collaborative filtering*, dan metode *k-means clustering*. Pada penelitian ini, sistem rekomendasi film pada platform *streaming* dibuat menggunakan metode *content-based filtering* karena sesuai dengan tujuan penelitian ini, yaitu membuat sistem rekomendasi film yang dapat merekomendasikan film sesuai dengan preferensi atau selera pengguna, dengan cara mencari film yang mirip dengan film kesukaan pengguna sehingga nantinya pengguna platform *streaming* dapat dengan cepat memilih film yang akan ditonton. Sistem rekomendasi sendiri merupakan sebuah metode yang digunakan untuk memberikan rekomendasi pada sebuah produk seperti buku, musik dan film dengan memberikan nilai prediksi tertinggi pada penggunaannya [3]. Sistem rekomendasi ini dibuat dengan mencari kemiripan dari satu *item* dengan *item* lainnya dan membantu dalam mengatasi informasi yang meluap dengan memberikan rekomendasi spesifik bagi pengguna dan diharapkan rekomendasi tersebut bisa memenuhi keinginan dan kebutuhan pengguna sehingga didapatkan sebuah rekomendasi yang dapat membantu pengguna [4].

Penerapan metode *Content-Based Filtering* dalam pembuatan sistem rekomendasi film pada platform *streaming* dapat menjadi solusi dikarenakan cara kerja metode ini, yaitu dengan membandingkan isi konten dari satu produk dengan produk lainnya untuk dicari kesamaannya. Hal ini dibuktikan dengan penelitian yang berhasil dilakukan oleh Putra dan Santika (2020) dalam pembuatan rekomendasi musik yang sesuai dengan preferensi pengguna menggunakan metode *Content-Based Filtering* [5]. Selain itu, pada penelitian terdahulu yang pernah dilakukan oleh Fajriansyah, Adikara, dan Widodo (2021) mengenai Sistem Rekomendasi Film Menggunakan *Content Based Filtering*, penelitian ini memberikan hasil akurasi pengujian sebesar 0,823254 dalam memberikan rekomendasi berdasarkan kemiripan film [6], hal-hal ini yang menjadi acuan peneliti dalam melakukan penelitian pengembangan sistem rekomendasi film pada platform *streaming* menggunakan metode yang sama. *Content-Based Filtering* sendiri adalah metode yang mempertimbangkan perilaku dari pengguna dari masa lalu yang kemudian diidentifikasi pola perilakunya untuk merekomendasikan barang yang sesuai dengan pola perilaku tersebut [7]. Metode *content-based filtering* akan mengekstrak informasi yang terdapat pada barang kemudian membandingkannya dengan informasi barang yang pernah dilihat atau disukai oleh pengguna [8]. Dalam penelitian ini, metode *Content-Based Filtering* diterapkan untuk mencari persamaan film melalui deskripsi film dengan memanfaatkan metode *TF-IDF* dan *Cosine Similarity* untuk melakukan pembobotan kata pada deskripsi film. Metode *TF-IDF* dan *Cosine Similarity* merupakan metode yang digunakan untuk melakukan pembobotan kata, dimana cara kerja *TF-IDF* adalah dengan memberikan bobot kepada kata yang paling sering keluar, sedangkan *Cosine Similarity* memberikan bobot kata dengan menghitung nilai kosinus sudut kata yang memiliki kesamaan,

dengan menggunakan konsep derajat kosinus yang hasilnya dibatasi antara 0 dan 1. Hasil dari pembobotan kata inilah yang nantinya akan digunakan untuk menyelesaikan tujuan dari penelitian ini, yaitu untuk menciptakan suatu sistem rekomendasi film pada platform *streaming* yang bisa membantu pengguna dalam menentukan pilihan film yang akan ditonton yang sesuai dengan preferensi selera mereka dan dalam waktu yang lebih cepat.

RESEARCH METHODS

Metodologi penelitian atau langkah-langkah yang dilalui selama penelitian diperlukan agar penelitian dapat berjalan dengan lancar dari awal hingga akhir penelitian. Berikut adalah metodologi penelitian yang dilakukan dalam penelitian sistem rekomendasi film pada platform *streaming* menggunakan metode *content-based filtering* dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

1. Data Collection

Data Collection atau pengumpulan data merupakan metode yang digunakan untuk menjangkau data yang telah ditentukan sesuai dengan kebutuhan penelitian [9]. Pengumpulan data merupakan kegiatan yang sangat penting karena tanpa data penelitian tidak akan tercapai tujuannya [10]. Dalam penelitian ini, teknik pengumpulan data yang digunakan adalah teknik

pengumpulan data sekunder karena data diperoleh tidak secara langsung dari objek penelitian [11], data pada penelitian ini diambil dari sumber yang telah tersedia, yaitu website Kaggle.com.

- 1) Data-data yang akan digunakan adalah dataset film-film yang ada/tersedia di beberapa platform streaming, contohnya seperti judul film, deskripsi film, dan genre film.
- 2) Dalam penelitian ini, dataset yang digunakan adalah dataset dari platform *streaming* Disney Hotstar, Netflix, Paramount, Apple TV, HBO, dan Amazon Prime Video, dataset ini dapat dilihat pada tautan berikut (<https://github.com/JeremiaMahes/DatasetSistemRekomendasiFilm>).

2. Data Preprocessing

Data *preprocessing* merupakan proses penting dalam pengolahan data yang dilakukan sebelum data dapat digunakan untuk analisis atau pemodelan. Tujuan utama dari tahap ini adalah untuk memastikan bahwa data yang akan digunakan untuk adalah akurat, lengkap, dan terstruktur dengan baik. Data yang salah dapat berdampak buruk terhadap data yang akan dianalisa dan dapat menghasilkan hasil yang tidak akurat [12]. Melihat bagaimana karakteristik data itu sendiri, pemilihan metode *preprocessing* apa saja yang akan digunakan dapat berpengaruh pada kualitas data keluaran proses. Pada penelitian ini, tahapan yang dilakukan dalam data *preprocessing* adalah melakukan pengecekan dan penanganan data duplikat serta melakukan pengecekan dan penanganan *missing values*. Dimana data duplikat adalah data yang memiliki nilai yang sama atau kembar dalam semua atribut sedangkan *missing values* adalah data-data yang bernilai kosong, kedua masalah ini harus ditangani supaya data dapat diolah dengan baik.

3. Split Data

Split data atau pemisahan data dalam penelitian ini bertujuan untuk memisahkan data film sesuai dengan tipenya, data-data film dipisahkan menjadi dua tipe yaitu tipe *movie* dan tipe *show*. Film dengan tipe *movie* adalah film yang tidak memiliki episode, contohnya adalah film 'The Silence', sedangkan film dengan tipe *show* adalah film yang mempunyai episode didalamnya, contohnya seperti drama Korea, anime, ataupun series Amerika. Pemisahan data ini dilakukan dengan tujuan supaya di dalam penelitian ini akan tercipta suatu sistem rekomendasi yang dapat merekomendasikan film bukan hanya dari kemiripannya saja, melainkan juga sesuai tipe film yang ada, sehingga nantinya hal ini dapat membantu pengguna dalam mencari rekomendasi film yang sesuai dengan kebutuhan pengguna.

4. Word Weighting

Penelitian ini melakukan *word weighting* atau pembobotan kata yang ada pada deskripsi film dengan menggunakan metode *tf-idf* dan juga *cosine similarity*. Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) merupakan metode untuk menghitung bobot suatu kata (*term*) terhadap dokumen [13], metode ini merupakan metode pembobotan term secara statistik yang banyak digunakan sebagai vektorisasi pada teks analisis [14]. Dalam penelitian ini, pembobotan *tf-idf* dihitung menggunakan *library python* yakni *scikit learn* dengan persamaan sebagai berikut:

$$\text{Rumus TF-IDF:}$$
$$\text{tf}(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

$$\text{idf}(t) = \log \frac{1+n}{1+\text{df}(t)} + 1 \quad (2)$$

$$\text{tfidf}(t, d) = \text{tf}(t, d) \times \text{idf}(t) \quad (3)$$

Keterangan:

$f_{t,d}$ = jumlah munculnya kata (*term*)

$\sum_{t' \in d} f_{t',d}$ = jumlah dokumen yang mengandung *term*

n = jumlah total dokumen pada dokumen set

$\text{df}(t)$ = jumlah dokumen di dalam dokumen set yang mengandung *term*

Selain *TF-IDF*, *cosine similarity* juga dapat digunakan untuk menghitung nilai kemiripan antara 2 item [15]. Setelah menghitung pembobotan kata menggunakan *TF-IDF*, langkah selanjutnya adalah menghitung hasil *TF-IDF* menggunakan *Cosine Similarity* sehingga nantinya pembobotan kata menghasilkan bobot yang maksimal untuk digunakan dalam pembuatan sistem rekomendasi film. Metode *Cosine Similarity* ini menghitung nilai kosinus sudut kata yang memiliki kesamaan, menggunakan konsep derajat kosinus yang hasilnya dibatasi antara 0 dan 1. Jika nilainya 0 maka dikatakan berbeda, namun apabila jika nilai hasil kesamaan kosinus adalah 1 maka dikatakan mirip [16]. Penghitungan *cosine similarity* pada penelitian ini juga menggunakan *library scikit learn* dengan formulanya adalah sebagai berikut:

Rumus Cosine Similarity:

$$k(x, y) = \frac{xy}{\|x\| \|y\|} \quad (4)$$

Keterangan:

x dan y = vektor matrix tf-idf

$\|x\|$ dan $\|y\|$ = norm dari x dan y

Pada penelitian ini, penghitungan pembobotan kata menggunakan *TF-IDF* dan *Cosine Similarity* dibantu dengan *library* bahasa pemrograman *Python*, yaitu dengan menggunakan *library Scikit-Learn*.

```
tfidf = TfidfVectorizer(stop_words='english')

tfidf_matrix_movies = tfidf.fit_transform(movies['description'])
tfidf_matrix_shows = tfidf.fit_transform(shows['description'])

cos_sim_movies = linear_kernel(tfidf_matrix_movies, tfidf_matrix_movies)
cos_sim_shows = linear_kernel(tfidf_matrix_shows, tfidf_matrix_shows)
```

Gambar 2. Pembobotan Kata dengan *TF-IDF* dan *Cosine Similarity*

RESULTS

Penelitian sistem rekomendasi film pada platform *streaming* menggunakan metode *content-based filtering* ini memanfaatkan *tf-idf* dan *cosine similarity* untuk melakukan pembobotan kata pada deskripsi film dari data-data film yang ada pada platform *streaming*, data-data ini diperoleh dari *website* Kaggle.com dan digabungkan menjadi satu kesatuan data seperti yang terlihat pada gambar di bawah.

Sistem Rekomendasi Film pada Platform Streaming Menggunakan Metode Content-Based Filtering (Jeremia Maheswara Azri Saputra)

id	title	type	description	release_year	age_certification	runtime	genres	production_countries	seasons	imdb_id	imdb_score	imdb_votes	tmdb_popularity	tmdb_score	
11321	tm57241	Never Been Kissed	MOVIE	Josie Geller, a baby-faced junior copywriter a...	1999	PG-13	107	['comedy', 'drama', 'romance']	[US]	NaN	tt0151738	6.0	93238.0	17.420	6.180
11387	ts22130	Rolie Polie Olie	SHOW	Rolie Polie Olie was a children's television s...	1998	TV-Y	21	['animation', 'comedy', 'family', 'fantasy', 'i...']	[CA, 'FR', 'US', 'GB']	6.0	tt0172049	6.3	3012.0	13.848	6.600
11586	tm98015	The Last Song	MOVIE	A drama centered on a rebellious girl who is s...	2010	PG	107	['drama', 'romance', 'music']	[US]	NaN	tt1294226	6.0	89378.0	15.081	7.242
11610	ts22233	Shake It Up	SHOW	Best pals CeCe and Rocky dream of dancing star...	2010	TV-G	25	['comedy', 'family']	[US]	3.0	tt0453993	8.0	88.0	41.672	7.800
11822	ts7273	Doc McStuffins	SHOW	A young African-American girl aspires to be a ...	2012	TV-G	22	['animation', 'family', 'fantasy', 'music']	[US]	6.0	tt1710295	6.6	2551.0	35.228	5.800
11955	tm152954	Zapped	MOVIE	Zoey is a talented dancer whose organized life...	2014	NaN	102	['comedy', 'family', 'fantasy']	[CA]	NaN	tt3544734	5.0	5749.0	12.426	6.100
12002	ts52840	Mars	SHOW	The maiden crew of the Daedalus spacecraft mus...	2016	TV-PG	47	['scifi', 'action', 'drama']	[US]	2.0	tt4939064	7.4	14872.0	13.596	7.010
12825	tm981436	Doc McStuffins: The Doc Is In	MOVIE	Doc McStuffins meets real kids, families, doct...	2020	NaN	24	['animation', 'family', 'fantasy']	[US]	NaN	tt13522660	NaN	NaN	1.084	7.000

Gambar 3. Data Film pada Platform Streaming

Data yang diperoleh berjumlah dua puluh lima ribu dua ratus empat puluh enam (25.246) data, namun ini masih data mentah yang nantinya akan melalui proses *preprocessing* data untuk menghandling data duplikat dan missing values, sehingga data akan siap untuk diproses ke tahap berikutnya.

Tabel 1. Missing values

Attribute	Sebelum Penanganan	Sesudah Penanganan
description	195	0
age_certification	11981	0
seasons	17739	0
imdb_id	1950	0
imdb_score	2558	0
imdb_votes	2618	0
tmdb_popularity	720	0
tmdb_score	3000	0

id	title	type	description	release_year	age_certification	runtime	genres	production_countries	seasons	imdb_id	imdb_score	imdb_votes	tmdb_popularity	tmdb_score	
0	tm87233	It's a Wonderful Life	MOVIE	A holiday favourite for generations... George...	1946	PG	130	['drama', 'family', 'fantasy', 'romance', 'com...']	[US]	NaN	tt0038650	8.6	467766.0	27.611	8.261
1	tm143047	Duck Soup	MOVIE	Rufus T. Firefly is named president/dictator o...	1933	NaN	69	['comedy', 'war']	[US]	NaN	tt0023969	7.8	60933.0	9.013	7.357
2	tm83884	His Girl Friday	MOVIE	Hildy, the journalist former wife of newspaper...	1940	NaN	92	['drama', 'romance', 'comedy']	[US]	NaN	tt0032599	7.8	60244.0	14.759	7.433
3	ts20945	The Three Stooges	SHOW	The Three Stooges were an American vaudeville ...	1934	TV-PG	19	['comedy', 'family']	[US]	26.0	tt0850645	8.5	1149.0	15.424	7.600
4	tm5012	Red River	MOVIE	Headstrong Thomas Dunson starts a thriving	1948	NaN	133	['western', 'drama', 'romance']	[US]	NaN	tt0040724	7.8	32210.0	12.400	7.400

Gambar 4. Data yang Memiliki Duplikat

Setelah mengatasi *missing values* dan data duplikat, jumlah data yang diperoleh berkurang menjadi delapan belas ribu tiga ratus tujuh puluh empat (18374) data, berikutnya adalah memisahkan data sesuai dengan tipenya, yaitu menjadi data film dengan tipe *movie* dan data dengan tipe *show*.

id	title	type	description	tahun_rilis	runtime	seasons	imd_score	imd_votes	tmdb_popularity	tmdb_score	production_country	genre	platform_streaming	rekomendasi_film	
0	tm87233	It's a Wonderful Life	MOVIE	A holiday favourite for generations... George...	1946	130	0.0	8.6	467766.0	27.611	8.261	US	drama	[amazon]	It's a Wonderful Life
1	tm143047	Duck Soup	MOVIE	Rufus T. Firefly is named president/dictator o...	1933	69	0.0	7.8	60933.0	9.013	7.357	US	comedy	[amazon]	Duck Soup
2	tm83884	His Girl Friday	MOVIE	Hildy, the journalist former wife of newspaper...	1940	92	0.0	7.8	60244.0	14.759	7.433	US	drama	[amazon, paramount+]	His Girl Friday
3	tm5012	Red River	MOVIE	Headstrong Thomas Dunson starts a thriving Tex...	1948	133	0.0	7.8	32210.0	12.400	7.400	US	western	[amazon, paramount+]	Red River
4	tm62253	The Best Years of Our Lives	MOVIE	It's the hope that sustains the spirit of ever...	1947	171	0.0	8.1	66209.0	16.056	7.838	US	drama	[amazon]	The Best Years of Our Lives

Gambar 5. Data Film Tipe *Movie*

id	title	type	description	tahun_rilis	runtime	seasons	imd_score	imd_votes	tmdb_popularity	tmdb_score	production_country	genre	platform_streaming	rekomendasi_film	
0	ts20945	The Three Stooges	SHOW	The Three Stooges were an American vaudeville ...	1934	19	26.0	8.5	1149.0	15.424	7.600	US	comedy	[amazon]	The Three Stooges
1	ts37076	The Jack Benny Program	SHOW	Laugh along with funnyman Jack Benny as he bri...	1950	30	21.0	8.6	1529.0	9.681	7.500	US	comedy	[amazon]	The Jack Benny Program
2	ts55748	What's My Line?	SHOW	Four panelists must determine guests' occupati...	1950	30	18.0	8.5	1658.0	45.113	7.200	US	reality	[amazon]	What's My Line?
3	ts3386	The Avengers	SHOW	The Avengers is a British television series cr...	1961	51	8.0	8.3	8195.0	42.586	7.853	GB	scifi	[amazon]	The Avengers
4	ts42867	Mister Rogers' Neighborhood	SHOW	Mister Rogers' Neighborhood is an American chi...	1968	28	31.0	8.7	9043.0	36.060	4.845	US	family	[amazon]	Mister Rogers' Neighborhood

Gambar 6. Data Film Tipe *Show*

Kemudian sesudah memisahkan data sesuai tipenya, langkah selanjutnya adalah memberi bobot kata pada deskripsi film menggunakan metode *tf-idf* dan *cosine similarity*, berikut adalah hasil dari pembobotan kata pada deskripsi film.

Tabel 2. Hasil Perhitungan TF-IDF Data Tipe *Movie*

(Indeks Doc, Indeks Term)	Nilai tf-idf
(0, 3229)	0.1315725588151985
(0, 32155)	0.17416978494031507
(0, 18563)	0.12432100515641953
(0, 1)	0.1436303044152082
(0, 4753)	0.12181721218786487
(0, 10726)	0.14675486790121026
.....
(13830, 29762)	0.2972132621957889
(13830, 9067)	0.32741121482597724
(13830, 11152)	0.21356582135494578
(13830, 18215)	0.12047235237518195

Tabel 3. Hasil Perhitungan TF-IDF Data Tipe *Show*

(Indeks Doc, Indeks Term)	Nilai tf-idf
(0, 3229)	0.1315725588151985
(0, 32155)	0.17416978494031507
(0, 18563)	0.12432100515641953
(0, 1)	0.1436303044152082
(0, 4753)	0.12181721218786487
(0, 10726)	0.14675486790121026
.....
(13830, 29762)	0.2972132621957889
(13830, 9067)	0.32741121482597724
(13830, 11152)	0.21356582135494578
(13830, 18215)	0.12047235237518195

Hasil dari perhitungan menggunakan *TF-IDF* akan digunakan untuk dihitung lagi menggunakan metode *cosine similarity* supaya hasil dari pembobotan kata menjadi semaksimal mungkin agar dapat memberikan rekomendasi yang sesuai dengan kemiripan film.

```
[[1.      0.      0.      ... 0.      0.      0.00846366]
 [0.      1.      0.      ... 0.      0.      0.      ]
 [0.      0.      1.      ... 0.01278883 0.      0.      ]
 ...
 [0.      0.      0.01278883 ... 1.      0.      0.      ]
 [0.      0.      0.      ... 0.      1.      0.      ]
 [0.00846366 0.      0.      ... 0.      0.      1.      ]]
```

Gambar 7. Hasil Perhitungan *Cosine Similarity* Data Tipe *Movie*

```
[[1.      0.      0.      ... 0.      0.      0.00720871]
 [0.      1.      0.      ... 0.      0.      0.02856115]
 [0.      0.      1.      ... 0.      0.      0.      ]
 ...
 [0.      0.      0.      ... 1.      0.      0.      ]
 [0.      0.      0.      ... 0.      1.      0.      ]
 [0.00720871 0.02856115 0.      ... 0.      0.      1.      ]]
```

Gambar 8. Hasil Perhitungan *Cosine Similarity* Data Tipe *Show*

Setelah data melewati berbagai proses pengolahan data, diantaranya yaitu *preprocessing* data, membagi pengelompokkan data sesuai dengan tipe film, dan pembobotan kata dari deskripsi film menggunakan metode *TF-IDF* dan *Cosine Similarity*, didapatkan hasil akhir berupa sebuah sistem yang dapat memberikan rekomendasi film yang mirip dan sesuai dengan film kesukaan/selera pengguna.

```
Judul: The Silence | Tahun: 2019
-----
rekomendasi_film  genre  tahun_rilis  platform_streaming
13682             A Quiet Place  horror        2018              [paramount+]
6220              A Quiet Place Part II  thriller      2021              [amazon, paramount+]
6540              All Creatures Here Below  drama        2022              [amazon]
6723              Fear of a Black Planet  thriller      2021              [amazon]
6029              Intolerance: No More  drama        2020              [amazon]
3001              Elfie Hopkins      horror        2012              [amazon, paramount+]
13620             10 Cloverfield Lane  horror        2016              [paramount+]
7584              Adventures in Babysitting  comedy       2016              [disney+]
11595             Svaha: The Sixth Finger  thriller      2019              [netflix]
3081              The Dark Lurking    horror        2011              [amazon]
-----
```

Gambar 9. Rekomendasi Film Tipe *Movie*

Gambar di atas merupakan hasil dari rekomendasi film dengan tipe *movie*, yaitu film ‘*The Silence*’ yang rilis di tahun 2019 dengan rekomendasinya diantaranya adalah film ‘*A Quiet Place*’, film ‘*A Quiet Place Part II*’, dan lain sebagainya seperti yang tertera di Gambar 9. Selain memberikan rekomendasi, sistem ini juga memberi info di platform mana film-film yang direkomendasikan tersebut dapat ditonton.

Judul: Stranger Things | Tahun: 2016

	rekendasi_film	genre	tahun_rilis	platform_streaming
1102	Home Before Dark	crime	2020	[appletv]
4540	Wolf Pack	drama	2023	[paramount+]
769	Hellier	reality	2019	[amazon]
1321	The Evermoor Chronicles	action	2014	[disney+]
1563	Secrets of Sulphur Springs	drama	2021	[disney+]
981	Suzhal - The Vortex	crime	2022	[amazon]
3161	The Society	drama	2019	[netflix]
2516	Dark	drama	2017	[netflix]
2046	Naomi	scifi	2022	[hbomax]
2722	Bitter Daisies	drama	2018	[netflix]

Gambar 10. Rekomendasi Film Tipe *Show*

Tidak hanya sistem rekomendasi film dengan tipe *movie*, namun sistem rekomendasi film dengan tipe *show* juga berhasil dibuat. Seperti yang terlihat pada Gambar 10, *show* dengan judul ‘*Stranger Things*’ yang dibuat pada tahun 2016 memiliki rekomendasi film yang mirip, yaitu film ‘*Home Before Dark*’, ‘*Wolf Pack*’, dan lainnya. Sistem ini juga memberikan info platform *streaming* mana yang memiliki koleksi film yang telah direkomendasikan sehingga para pengguna tidak akan kebingungan lagi dalam mencari rekomendasi film dan tempat atau platform menontonnya.

DISCUSSION

Hasil rekomendasi dari sistem rekomendasi film pada platform *streaming* selanjutnya akan di evaluasi menggunakan 3 metode, yaitu *Precision@K* ($P@K$), *Average Precision@K* ($AP@K$), dan *Mean Average Precision@K* ($MAP@K$). Hasil rekomendasi film di evaluasi dengan mempertimbangkan jalan cerita atau plot film dan juga genre film, pengujian hasil rekomendasi film ini melibatkan 3 partisipan untuk menguji masing-masing 3 film untuk setiap hasil sistem rekomendasi film pada platform *streaming*, baik itu hasil dari sistem rekomendasi film dengan tipe *movie* maupun hasil dari sistem rekomendasi film dengan tipe *show*.

Tabel 4. Hasil Perhitungan $P@K$ dan $AP@K$ Sistem Rekomendasi Film Tipe *Movie*

Partisipan	Judul	Peringkat	$P@K$	$AP@K$	Rata-rata
Partisipan 1	The Silence	1	1	0.728	0.73
		∴	∴		
	Frozen	10	0.76	0.672	
		∴	∴		
	Avengers: Age of Ultron	10	0.5	0.79	
		∴	∴		
Partisipan 2	The Fast and the Furious	10	0.5	0.722	0.674
		∴	∴		
	Cars	10	1	0.585	
		∴	∴		

Partisipan 3	Superman	10	0.33	0.715	0.70266 66667
		1	1		
		∴	∴		
	Justice League	10	0.43	0.7	
		1	1		
		∴	∴		
	Toy Story	10	0.48	0.688	
		1	1		
		∴	∴		
	Percy Jackson & the Olympians: The Lightning Thief	10	0.51	0.72	
		1	1		
		∴	∴		
		10	0.6		

Tabel 5. Hasil Perhitungan P@K dan AP@K Sistem Rekomendasi Film Tipe Show

Partisipan	Judul	Peringkat	P@K	AP@K	Rata-rata
Partisipan 1	Stranger Things	1	1	0.856	0.783
		∴	∴		
		10	0.67		
	Power Rangers Ninja Steel	1	1	0.793	
		∴	∴		
		10	0.88		
	Naruto	1	0.9	0.7	
		∴	∴		
		10	0.63		
Partisipan 2	Oh My Ghost	1	0.9	0.782	0.82366 66667
		∴	∴		
		10	0.62		
	Welcome to Waikiki	1	0.86	0.862	
		∴	∴		
		10	0.43		
	Bluey	1	1	0.827	
		∴	∴		
		10	0.9		
Partisipan 3	The Walking Dead	1	0.72	0.801	0.78533 33
		∴	∴		
		10	0.44		
	Wednesday	1	1	0.748	
		∴	∴		
		10	0.56		
	Bring it On, Ghost	1	1	0.807	
		∴	∴		
		10			

10 0.85

Perhitungan $P@K$ dilihat dari relevansi hasil rekomendasi film dengan film yang dicari rekomendasinya, jika hasil rekomendasi film relevan dengan film yang dicari rekomendasinya maka nilai dari $P@K$ adalah 1, dan jika tidak relevan akan dinilai kurang dari 1 dengan mempertimbangkan kesamaan plot film dan juga genre film. Setelah mendapatkan nilai $P@K$ langkah selanjutnya adalah menghitung nilai $AP@K$ dengan membuat rata-rata nilai $P@K$.

Tabel 6. Nilai $MAP@K$ Pada Sistem Rekomendasi Film Pada Platform *Streaming*

Sistem Rekomendasi Film	Nilai $MAP@K$
Tipe <i>Movie</i>	0.702222223
Tipe <i>Show</i>	0.797333322

Hasil evaluasi sistem rekomendasi film pada platform *streaming* dapat dilihat pada nilai $MAP@K$ yang telah diperoleh dengan menghitung *mean* dari nilai $AP@K$ yang telah didapatkan. Hasil dari evaluasi sistem rekomendasi film pada platform *streaming* menunjukkan performa terbaik sistem ini ada pada sistem rekomendasi film tipe *show* yaitu dengan skor 0.7973 sedangkan untuk sistem rekomendasi film tipe *movie* memperoleh skor akurasi sebesar 0.7022.

CONCLUSIONS AND RECOMMENDATIONS

Penelitian ini menerapkan metode *Content-Based Filtering* dalam pembuatan sebuah sistem rekomendasi film pada platform *streaming*, metode ini digunakan untuk mencari kesamaan-kesamaan film dari deskripsi filmnya dengan memberikan pembobotan kata pada deskripsi film. Pembobotan dilakukan dengan menghitung bobot dari atribut deskripsi menggunakan metode *TF-IDF* dan *Cosine Similarity*, dengan membagi film berdasarkan tipenya, yaitu data dengan tipe *movie* dan data dengan tipe *show*, sebuah sistem rekomendasi film pada platform *streaming* berhasil dibuat. Dengan membagi film sesuai tipenya, pengguna akan mendapatkan rekomendasi film yang lebih akurat, karena bukan hanya membuat rekomendasi berdasarkan kemiripan film saja, namun juga rekomendasi sesuai dengan tipe film. Skor evaluasi sistem rekomendasi film pada platform *streaming* menggunakan metode *content-based filtering* ini adalah 0.7973 untuk sistem rekomendasi film tipe *show* dan 0.7022 untuk sistem rekomendasi film tipe *movie*.

Dari kesimpulan penelitian ini, rekomendasi untuk penelitian berikutnya adalah dengan menggunakan dataset yang lebih kaya dan beragam, seperti menambahkan data dari platform *streaming* lain yang belum ada pada penelitian ini. Selain itu, rekomendasi untuk penelitian selanjutnya adalah dengan menambahkan fitur-fitur *custom* atau atribut tambahan yang dapat meningkatkan akurasi rekomendasi, misalnya mempertimbangkan ulasan pengguna, aktor favorit, atau faktor lain yang dapat mempengaruhi preferensi pengguna.

REFERENCES

- [1] F. T. Laily and A. P. Purbantina, "Digitalisasi Industri Perfilman Korea Selatan Melalui Netflix Sebagai Alternatif Pasar Ekspor Film," *Expose: Jurnal Ilmu Komunikasi*, vol. 4, no. 2, p. 141, Dec. 2021, doi: 10.33021/exp.v4i2.1494.
- [2] V. Sandrya and D. Arisandi, "SISTEM REKOMENDASI FILM MENGGUNAKAN METODE MULTIPLE ATTRIBUTE UTILITY THEORY," 2022.

- [3] K. Ratna Sari, W. Suharso, and Y. Azhar, "Pembuatan Sistem Rekomendasi Film dengan Menggunakan Metode Item Based Collaborative Filtering pada Apache Mahout," *REPOSITOR*, vol. 2, no. 6, pp. 767–774, 2020.
- [4] Y. Visher Laja Jaja, B. Susanto, L. Ricky Sasongko, and K. Kunci, "Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens," 2020. [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/decartesian>
- [5] A. I. Putra and R. R. Santika, "Implementasi Machine Learning dalam Penentuan Rekomendasi Musik dengan Metode Content-Based Filtering," *Edumatic : Jurnal Pendidikan Informatika*, vol. 4, no. 1, pp. 121–130, Jun. 2020, doi: 10.29408/edumatic.v4i1.2162.
- [6] M. Fajriansyah, P. P. Adikara, and A. W. Widodo, "Sistem Rekomendasi Film Menggunakan Content Based Filtering," 2021. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [7] S. Reddy, S. Nalluri, S. Kuniseti, S. Ashok, and B. Venkatesh, "Content-based movie recommendation system using genre correlation," in *Smart Innovation, Systems and Technologies*, Springer Science and Business Media Deutschland GmbH, 2019, pp. 391–397. doi: 10.1007/978-981-13-1927-3_42.
- [8] Fatoni Batari Agung Larasati and Herny Februariyanti, "SISTEM REKOMENDASI PRODUCT EMINA COSMETICS DENGAN MENGGUNAKAN METODE CONTENT - BASED FILTERING," *MISI (Jurnal Manajemen informatika & Sistem Informasi)*, vol. 4, pp. 45–54, Jan. 2021.
- [9] P. Darasa Panjaitan, E. Purba, and D. Damanik, "PENGARUH JUMLAH UANG BEREDAR DAN NILAI TUKAR TERHADAP INFLASI DI SUMATERA UTARA," *EKUILNOMI : Jurnal Ekonomi Pembangunan*, vol. 3, no. 1, pp. 2614–7181, 2021, doi: 10.36985/ekuilnomi.v3i1.76.
- [10] I. Firdaus, R. Hidayati, R. S. Hamidah, R. Rianti, R. Cahyuni, and K. Khotimah, "Model-Model Pengumpulan Data dalam Penelitian Tindakan Kelas," *Jurnal Kreativitas Mahasiswa*, vol. 1, no. 2, pp. 105–113, 2023.
- [11] Baharuddin, Masnur, and Rismayani, "APLIKASI VIRTUAL TOUR FAKULTAS TEKNIK BERBASIS ANDROID MOBILE," *JURNAL SINTAKS LOGIKA (JSilog)*, vol. 1, no. 2, pp. 116–122, 2021, doi: 10.31850/jsilog.v1i3.
- [12] I. Hadi, L. W. Santoso, and A. N. Tjondrowiguno, "Sistem Rekomendasi Film menggunakan User-based Collaborative Filtering dan K-modes Clustering," 2020.
- [13] A. Riyani, M. Zidny Naf'an #2, and A. Burhanuddin, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," 2019.
- [14] B. Bayu Baskoro, I. Susanto, and S. Khomsah, "Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)," *Journal of Informatics, Information System, Software Engineering and Applications*, vol. 3, no. 2, pp. 21–29, 2021, doi: 10.20895/INISTA.V3I2.
- [15] S. A. Zulvian, K. Prihandani, and A. A. Ridha, "PERBANDINGAN METODE MSD DAN COSINE SIMILARITY PADA SISTEM REKOMENDASI ITEM-BASED COLLABORATIVE FILTERING," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 4, no. 2, 2021, [Online]. Available: <https://nijianmo.github.io/>.
- [16] K. Telaumbanua and L. Nababan, "Jurnal IEED (Informatics Engineering and Electronic Data) Implementasi Metode Cosine Similarity Dalam Mendeteksi Kemiripan Dan Perbedaan Gambar Hasil Scan Berbasis Android Article Info," Online, 2022.