

# Sistem Rekomendasi Film Menggunakan Content Based Filtering

Rizka Amelia Sari<sup>1, a)</sup>, Septi Fajar Isnaini<sup>2, b)</sup>, Erni Seniwati<sup>3, c)</sup>

<sup>1,2,3)</sup>Program Studi Sistem Informasi, Universitas Amikom Yogyakarta, Jl. Padjajaran, Ring Road Utara, Yogyakarta 55283, Indonesia

Author Emails

<sup>a)</sup> [rizka30@students.amikom.ac.id](mailto:rizka30@students.amikom.ac.id)

<sup>b)</sup> [septiafajar@students.amikom.ac.id](mailto:septiafajar@students.amikom.ac.id)

<sup>c)</sup> Corresponding author: [erni.s@amikom.ac.id](mailto:erni.s@amikom.ac.id)

**Abstract.** A recommendation system is a technique used in technological developments to make it easier to make choices based on the recommendations provided. The recommendation system can be applied to various areas of data objects, including tourism, culinary, books, e-commerce, fashion, films, music, and others. In this research, a recommendation system was created for film recommendations using the content-based concept. The dataset used comes from Kaggle, namely <https://www.kaggle.com/datasets/georgescutelnicu/top-100-popular-movies-from-2003-to-2022-imdb>. The amount of film data used is 2000 film data. There are 13 variables used, namely title, rating, year, month, certificate, runtime, directors, stars, genre, filming location, budget, income and country of origin. The resulting recommendations will provide a sequence of recommendations based on similarity values. The recommendation similarity value uses the cosine similarity method. The results of implementation on the web by selecting the film title "Avengers: Infinity Wars" produced the highest similarity score for the film content, namely 97.53% with the film title "Avengers: Endgame". The recommendation system created can be used by film fans to help choose films whose content is similar to their favorite films.

**Keywords :** Content-based, Cosine Similarity, Movies, Similarity Values, Recommendation Systems

**Abstraksi.** Sistem rekomendasi merupakan suatu teknik yang digunakan pada perkembangan teknologi untuk memberikan kemudahan dalam menentukan pilihan berdasarkan rekomendasi yang diberikan. Sistem rekomendasi dapat diterapkan pada berbagai area objek data diantaranya pada area wisata, kuliner, buku, e-commerce, fashion, film, musik, dan lainnya. Pada penelitian ini sistem rekomendasi yang dibuat untuk rekomendasi film dengan menggunakan konsep content-based. Dataset yang digunakan berasal dari kaggle yaitu <https://www.kaggle.com/datasets/georgescutelnicu/top-100-popular-movies-from-2003-to-2022-imdb>. Jumlah data film yang digunakan 2000 data film. Variabel yang digunakan ada 13 yaitu title, rating, year, month, certificate, runtime, directors, stars, genre, filming location, budget, income, dan country of origin. Rekomendasi yang dihasilkan akan memberikan urutan rekomendasi berdasarkan nilai kesamaan. Nilai kesamaan rekomendasi menggunakan metode cosine similarity. Hasil implementasi pada web dengan memilih judul film "Avengers: Infinity Wars" menghasilkan nilai kesamaan tertinggi kemiripan content film nya yaitu 97,53% dengan judul film "Avengers: Endgame". Sistem rekomendasi yang dibuat dapat digunakan oleh penggemar film untuk membantu memilih film yang content nya mirip dengan film yang digemari.

**Kata Kunci :** Content-based, Cosine Similarity, Film, Nilai Kemiripan, Sistem Rekomendasi

## PENDAHULUAN

Film merupakan salah satu bentuk karya seni dan budaya yang digunakan sebagai alat ekspresi kreativitas serta media penyampaian pesan, nilai-nilai, dan hiburan kepada penonton. Secara umum, film terdiri dari dua unsur utama, yaitu unsur naratif yang berkaitan dengan cerita dan alur, serta unsur sinematik yang mencakup teknik pengambilan gambar, editing, suara, dan penyusunan adegan. Kedua unsur ini saling berinteraksi dan melengkapi untuk menciptakan sebuah film yang utuh. Film memiliki peran penting dalam kehidupan pengguna, baik sebagai hiburan, sarana edukasi, maupun media penyampaian pesan. Namun, dengan banyaknya film yang tersedia, pengguna seringkali kesulitan menemukan film yang sesuai dengan preferensi mereka. Hal ini menimbulkan masalah dalam hal

efisiensi waktu dan kepuasan pengguna. Oleh karena itu, diperlukan sistem rekomendasi yang dapat membantu pengguna menemukan film yang relevan berdasarkan preferensi mereka, seperti genre, sutradara, atau tema cerita.

Salah satu pendekatan yang efektif untuk mengatasi masalah ini adalah content-based filtering. Metode ini bekerja dengan menganalisis karakteristik atau fitur dari film yang telah dinilai atau dipilih oleh pengguna, kemudian mencocokkan film-film lain yang memiliki kesamaan fitur. Fitur-fitur tersebut dapat mencakup genre, sutradara, aktor, atau tema cerita. Untuk mengukur tingkat kesamaan antara film, algoritma cosine similarity sering digunakan. Algoritma ini menghitung kemiripan antara dua film berdasarkan sudut antara vektor fitur mereka. Semakin kecil sudutnya, semakin tinggi tingkat kesamaan antara film tersebut. Dengan menggunakan algoritma ini, sistem rekomendasi dapat memberikan rekomendasi yang lebih akurat dan relevan, sehingga pengguna dapat menemukan film yang sesuai dengan selera mereka. Dalam kehidupan nyata, konsep ini berperan penting dalam membantu pengguna menghemat waktu pencarian dan meningkatkan kepuasan mereka dalam memilih film.

Keterkaitan antara konsep sistem rekomendasi berbasis content-based filtering dengan masalah yang dihadapi pengguna terletak pada kemampuannya untuk menyederhanakan proses pencarian film. Dengan menerapkan algoritma cosine similarity, sistem dapat menganalisis kesamaan fitur antara film yang telah dinilai pengguna dengan film lainnya, sehingga memberikan rekomendasi yang lebih personal dan sesuai dengan preferensi pengguna. Tujuan dari penerapan sistem ini adalah untuk meningkatkan akurasi rekomendasi, menghemat waktu pencarian, dan memberikan pengalaman menonton yang lebih menyenangkan. Manfaat yang diharapkan adalah pengguna dapat menemukan film yang sesuai dengan selera mereka secara efisien, sehingga meningkatkan kualitas pengalaman menonton dan kepuasan secara keseluruhan.

## TINJAUAN PUSTAKA

Penelitian dengan topik sistem rekomendasi masih menjadi topik penelitian yang menarik. Menariknya topik penelitian ini dikarenakan kefleksibelan sistem rekomendasi yang dapat diterapkan di berbagai bidang atau objek. Bidang atau objek yang dapat diimplementasikan konsep sistem rekomendasi diantaranya adalah film, musik, wisata, kuliner, fashion dan masih banyak lagi. Konsep sistem rekomendasi terbagi beberapa bagian yaitu demografi, content-based, collaborative, dan hybrid filtering. Telah dilakukan penelitian yaitu sistem rekomendasi dengan konsep content-based yang diterapkan pada bidang pendidikan dimana sistem rekomendasi digunakan untuk penentuan film dengan studi kasus pada mata pelajaran data mining di SMK [1] [10] selain itu digunakan juga untuk e-commerce yaitu Shopee [3]. Selain itu terdapat penerapan collaborative filtering [2] untuk rekomendasi film pada IMDB. Hasil rekomendasi akan dipengaruhi oleh nilai kesamaan yang dapat memanfaatkan metode TF-IDF [9] atau pun metode lain nya yang mengadopsi dari metode-metode yang ada pada machine learning seperti decision tree, convolutional neural network, k-nearest neighbour [4][7][8] untuk menghasilkan validasi dari kinerja model yang digunakan. Selain penggunaan konsep content-based, terdapat juga penelitian lainnya yang menggunakan konsep hybrid filtering [6] bahkan sampai menggunakan multiple atribut [5]. Berdasarkan dari penelitian-penelitian sebelumnya yang sudah dilakukan oleh peneliti maka pada penelitian ini dibuat sistem rekomendasi film yang menggunakan metode TF-IDF dan cosine similarity. Metode TF-IDF dan cosine similarity digunakan untuk memberikan nilai kemiripan yang diwakilkan dengan memberikan urutan film yang direkomendasikan kepada penggemar film berdasarkan nilai kemiripan yang paling besar. Sehingga penggemar film dapat menikmati film dengan nilai kesukaan yang sangat mirip.

## METODE PENELITIAN

### Teori

- a. Content based filtering  
Content based filtering yaitu metode yang digunakan pada system rekomendasi yang berfokus pada karakteristik atau konten dari item-item yang ingin direkomendasikan atau dianalisis.
- b. TF-IDF (*Term Frequency-Inverse Document Frequency*)  
TF- IDF adalah metode statistik yang digunakan dalam pemrosesan bahasa alami dan sistem temu kembali informasi untuk mengukur pentingnya sebuah term (kata kunci) dalam sebuah dokumen dalam konteks koleksi dokumen yang lebih besar. Persamaan untuk perhitungan dari tf – idf terdapat pada persamaan (1).

$$TF - IDF(t, d) = TF(t, d) \times \log\left(\frac{N}{DF(t)}\right) \quad (1)$$

- c. *Cosine Similarity* adalah metode untuk mengukur kesamaan antara dua vektor dalam ruang multidimensi. Pada rekomendasi film, hal ini membantu menentukan seberapa mirip film-film berdasarkan fitur mereka. Rumus cosine similarity dapat dilihat pada persamaan (2).

$$\text{Cosine Similarity}(A,B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (2)$$

- d. *DF (Document Frequency)*

Df yaitu jumlah kemunculan sebuah term pada dokumen. Rumus document frequency dapat dilihat pada Persamaan (3).

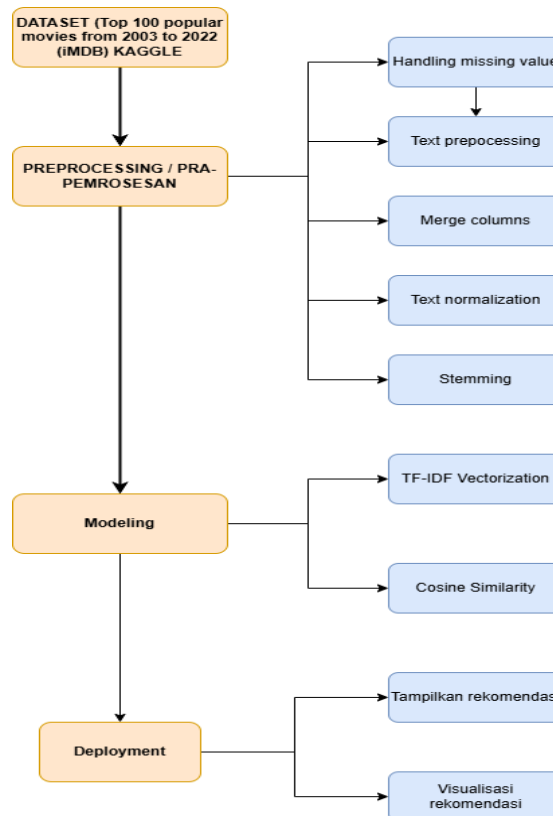
$$DF(t) = \text{Jumlah dokumen yang mengandung term } t \quad (3)$$

- e. *TF (Term Frequency)*

Tf dapat mengukur seberapa sering suatu kata muncul dalam sebuah dokumen. Persamaan (4) merupakan rumus term frequency.

$$TF(t, d) = \frac{\text{Jumlah kemunculan term } t \text{ dalam dokumen } d}{\text{Jumlah total term dalam dokumen } d} \quad (4)$$

## Alur Penelitian



GAMBAR 1. Alur Penelitian

Dari Gambar 1 terdapat 4 tahapan pada penelitian ini yaitu :

1. Pengumpulan Data

Pada penelitian ini menggunakan sekumpulan data yang dinamakan dataset. Dataset diperoleh dari Kaggle

yaitu <https://www.kaggle.com/datasets/georgescutelnicu/top-100-popular-movies-from-2003-to-2022-imdb>. Dataset ini merupakan kumpulan data film-film populer yang dirilis selama periode 20 tahun, yaitu dari tahun 2003 hingga 2022. Pada dataset tersebut terdapat 13 kolom yang berisikan title, rating, year, month, certificate, runtime, directors, stars, genre, filming location, budget, income dan country of origin. Secara keseluruhan, dataset ini berisi 2.000 data film.

2. Pre Processing

Drop Column. Menghapus kolom yang tidak diperlukan yaitu Month, Filming\_location, Budget, dan Income

a. Handling Missing Value

1. Terdapat missing value pada kolom Rating sebanyak 1 value dan pada kolom Certificate sebanyak 34 value, sehingga dilakukan penanganan sebagai berikut:
2. Mengisi missing value pada kolom Rating dengan nilai rata-rata
3. Mengisi missing value pada kolom Certificate dengan “Unknown”

b. Text Preprocessing

1. Mengganti koma dengan spasi pada kolom Directors, Stars, Genre, Country\_of\_origin
2. Untuk mengubah data yang belum terstruktur menjadi terstruktur.

c. Merge Columns. Menggabungkan kolom Certificate, Directors, Stars, Genre, dan Country\_of\_origin menjadi satu string pada kolom Tags

d. Text Normalization. Mengubah seluruh teks menjadi huruf kecil

e. Stemming. Tahap ini dilakukan untuk mengubah sebuah kata ke dalam bentuk dasarnya. contohnya seperti dari (adventure) menjadi (adventure)

3. Modelling

Penelitian ini menerapkan model yaitu pemanfaatan metode cosine similarity yang digunakan untuk menentukan nilai kesamaan dari satu film dengan film lain nya.

4. Deployment

Implementasi dari modelling pada penelitian ini menggunakan bahasa pemrograman Python dan memanfaatkan streamlit sebagai media web nya.

## HASIL DAN PEMBAHASAN

### Dataset

Title	Rating	Year	Month	Certificate	Runtime	Directors	Stars	Genre	Filming_loc	Budget	Income	Country_of_origin
Avatar: The	7.8	2022	December	PG-13	192	James Cameron	Sam Worthington	Action, Adv	New Zealand	\$350,000,000	\$2,267,946,983	United States
Guillermo c	7.6	2022	December	PG	117	Guillermo del Toro	Ewan McGregor, Jim Carrey	Animation, USA		\$35,000,000	\$108,967	United States, Mexico, France
Bullet Train	7.3	2022	August	R	127	David Leitch	Brad Pitt, Jason Statham	Action, Thr	Japan	\$85,900,000	\$239,268,602	Japan, United States
The Banshe	7.8	2022	November	R	114	Martin McDonnell	Colin Farrell, Saoirse Ronan	Comedy, D	Ireland	Unknown	\$19,720,823	Ireland, United Kingdom, United States
M3gan	6.4	2022	December	PG-13	102	Gerard John-Jones	Jenna Ortega, Allison Janney	Horror, Sci-Fi	New Zealand	\$12,000,000	\$171,253,910	United States
Emancipati	6.1	2022	December	R	132	Antoine Fuqua	Will Smith, Martin Lawrence	Action, Thr	Unknown	\$120,000,000	Unknown	United States
Amsterdam	6.1	2022	October	R	134	David O. Russell	Christian Bale, Brad Pitt, Robert Pattinson	Comedy, D	USA	\$80,000,000	\$31,245,810	United States, Japan
Violent Nig	6.9	2022	December	R	112	Tommy Wirkkunen	David Harbour, Jesse Plemons	Action, Cor	Canada	\$20,000,000	\$59,595,460	United States, Canada
The Whale	7.8	2022	December	R	117	Darren Aronofsky	Brendan Fraser, Hong Kong	Drama	USA	\$10,000,000	\$31,104,312	United States
The Fabelm	7.6	2022	November	PG-13	151	Steven Spielberg	Michelle Williams, Tom Holland	Drama	USA	\$40,000,000	\$9,500,361	United States
The Menu	7.5	2022	November	R	107	Mark Mylod	Ralph Fiennes, Annette Bening	Comedy, H	USA	\$35,000,000	\$65,878,071	United States
Babylon	7.3	2022	December	R	188	Damien Chazelle	Brad Pitt, Dianna Agron	Comedy, D	USA	\$78,000,000	\$61,851,455	United States
X	6.6	2022	March	R	105	Ti West	Mia Goth, John Wood	Horror, My	New Zealand	\$1,000,000	\$14,779,858	United States, Canada
Bones and	7	2022	November	R	131	Luca Guadagnino	Timothée Chalamet, Saoirse Ronan	Drama, H	USA	\$16,000,000	\$14,134,907	Italy, United States
Black Adan	6.5	2022	October	PG-13	125	Jaume Collet-Sierralta	Wayne Newton, Joel Kinnaman	Action, Adv	USA	\$195,000,000	\$391,273,355	United States, Canada, New Zealand, Hungary
Spirited	6.6	2022	November	PG-13	127	Sean Anders	Will Ferrell, John Mulaney	Comedy, F	USA	\$100,000,000	Unknown	United States
Lady Chattr	6.7	2022	December	R	126	Laure de Clermont-Tonnerre	Emma Corrin, David Tennant	Drama, Rom	UK	Unknown	Unknown	United Kingdom, United States
A Christma	6.8	2022	November	PG	98	Clay Kaytis	Peter Billingsley, Peter Billingsley	Comedy, F	USA	Unknown	Unknown	United States, Mexico, Hungary, Canada
Troll	5.8	2022	December	TV-14	101	Roar Uthaug	Ine Marie Vrieland	Action, Adv	Norway	Unknown	Unknown	Norway
Black Pantf	6.8	2022	November	PG-13	161	Ryan Coogler	Letitia Wright, Ice Cube	Action, Adv	USA	\$250,000,000	\$791,852,924	United States
Glass Onio	7.6	2022	December	PG-13	140	Rian Johnson	Daniel Craig, Rami Malek	Comedy, Cr	Greece	\$40,000,000	\$13,280,000	United States
Triangle of	7.6	2022	September	R	147	Ruben Östlund	Thobias Thoenes, Lena Dunham	Comedy, D	Greece	€ 10,000,000	\$17,211,145	Sweden, France, United Kingdom, Germany, Turkey, Gr
Everything	8.1	2022	March	R	139	Dan Kwan	Michelle Yeoh, Daniel Craig	Action, Adv	USA	\$25,000,000	\$103,384,361	United States
Emily the C	6.7	2022	August	R	97	John Patton Aubrey	Plai Cosette, John Wood	Crime, D	Mexico	\$3,000,000	\$2,156,296	United States
God's Crool	7	2022	December	TV-MA	154	Oriol Paulo	Bárbara Lennie	Mystery	Spain	Unknown	\$6,060,070	Spain
Don't Worr	6.2	2022	September	R	123	Olivia Wilde	Florence Piau, David Oyelowo	Drama, Thr	USA	\$20,000,000	\$86,709,403	United States
Pinocchio	5.1	2022	September	PG	105	Robert Zemeckis	Joseph Gordon-Levitt, Tom Hanks	Adventure, I	Italy	\$150,000,000	\$37,353	United States

GAMBAR 2. Dataset Film

Dataset dapat diakses pada kaggle dengan judul [Top 100 Popular Movies from 2003 to 2022-IMDb].

## Konsep Sistem Rekomendasi

Cara Kerja Konsep Sistem Rekomendasi diuraikan berikut:

- a. Fitur Data yang Diperlukan  
 Pada proyek ini, digunakan fitur-fitur berikut dari dataset:
  1. Title: Judul film
  2. Certificate: Klasifikasi film
  3. Directors: Nama sutradara
  4. Stars: Nama bintang film
  5. Genre: Genre film
  6. Country\_of\_origin: Negara asal film
- b. Sample Data

TABEL 1. Tabel Sampel Data

Output	Title	Tags
1.	Avatar: The Way of Water	pg-13 james cameron sam worthington zoe saldana sigourney weaver stephen lang action adventure fantasy united states
2.	Guillermo del Toro's Pinocchio	pg guillermo del toro mark gustafson ewan mcgregor david bradley gregory mann burn gorman animation drama family united states mexico france
3.	Bullet Train	r david leitch brad pitt joey king aaron taylor johnson brian tyree henry action comedy thriller japan united states
4.	The Banshees of Inisherin	r martin mcdonagh colin farrell brendan gleeson kerry condon pat shortt comedy drama ireland united kingdom united states
5.	M3gan	pg-13 gerard johnstone jenna davis amie donald allison williams violet mcgraw horror sci-fi thriller united states

Langkah Perhitungan :

### 1. TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) merupakan metode pembobotan kata dalam analisis teks yang membantu menentukan pentingnya sebuah kata dalam dokumen.

#### a. TF (Term Frequency)

Pada langkah ini dilakukan perhitungan frekuensi kemunculan setiap kata kunci dalam dokumen. Sebagai contoh, kita ambil beberapa kata kunci dari Tags untuk perhitungan. Menghitung dengan beberapa kata kunci berikut : Action, adventure, comedy, pg, r, united.

Contoh Perhitungan TF (menggunakan persamaan (4))

Film A: Avatar: The Way of Water dengan total kata: 16

$$TF(\text{action}, A) = \frac{1}{16} = 0,0625$$

$$TF(\text{adventure}, A) = \frac{1}{16} = 0,0625$$

$$TF(\text{comedy}, A) = \frac{0}{16} = 0$$

$$TF(\text{pg}, A) = \frac{1}{16} = 0,0625$$

$$TF(\text{r}, A) = \frac{0}{16} = 0$$

$$TF(\text{united}, A) = \frac{1}{16} = 0,0625$$

Dicontohkan perhitungan TF untuk film A dan dilakukan proses perhitungan yang sama terhadap film B (Guillermo del Toro's Pinocchio) dengan jumlah kata yaitu 21, film C (Bullet Train) dengan jumlah kata yaitu 19, film D (The Banshees of Inisherin) dengan jumlah kata yaitu 18, dan film E (M3gan) dengan jumlah kata yaitu 16. Nilai TF(action, A), TF(adventure, A), TF(comedy, A), TF(pg, A), TF(r, A), dan TF(united, A) diperoleh dari jumlah kata yang ada pada Tabel 1 di film A dengan kolom Tags yaitu ada 16 kata. Kata action yang ada pada kolom Tags berjumlah 1 (film Avatar: The Way of Water)

yang kemudian dilakukan proses yang sama untuk penentuan jumlah kata pada kata-kata lain nya.

**b. DF (Document Frequency)**

Pada tahap ini dilakukan perhitungan jumlah film yang mengandung setiap kata kunci. Contoh perhitungan DF dilakukan dengan menggunakan persamaan (3) diperoleh hasil seperti berikut:

$$\begin{aligned} DF(action) &= 2 \\ DF(adventure) &= 1 \\ DF(comedy) &= 3 \\ DF(pg) &= 1 \\ DF(r) &= 3 \\ DF(united) &= 5 \end{aligned}$$

Nilai-nilai DF diperoleh dari data yang ada pada Tabel 1 kolom Tags. Terdapat 5 judul film dan masing-masing judul film memiliki uraian penanda masing-masing judul film. Dari semua uraian penanda di 5 judul film tersebut di hitung total kata kunci. Kata action berjumlah 2, kata adventure berjumlah 1, kata comedy berjumlah 3, kata pg berjumlah 1, kata r berjumlah 3, dan kata united berjumlah 5.

**c. TF-IDF**

TF-IDF merupakan perkalian dari TF dan IDF yang menggunakan persamaan (1).

Contoh perhitungan TF-IDF pada film A:

Misal data film yang digunakan ada 5 film (berdasarkan data di Tabel 1) maka nilai  $N=5$

$$\begin{aligned} TF - IDF(action, A) &= 0,0625 \times \log\left(\frac{5}{2}\right) = 0,0573 \\ TF - IDF(adventure, A) &= 0,0625 \times \log\left(\frac{5}{1}\right) = 0,1006 \\ TF - IDF(comedy, A) &= 0 \times \log\left(\frac{5}{3}\right) = 0 \\ TF - IDF(pg, A) &= 0,0625 \times \log\left(\frac{5}{1}\right) = 0,1006 \\ TF - IDF(r, A) &= 0 \times \log\left(\frac{5}{3}\right) = 0 \\ TF - IDF(united, A) &= 0,0625 \times \log\left(\frac{5}{5}\right) = 0 \end{aligned}$$

**2. Cosine Similarity**

Setelah mendapatkan representasi TF-IDF untuk setiap film, kita bisa menghitung cosine similarity antara Film A dan Film B (menggunakan persamaan (2)).

**Vektor TF-IDF:**

Film A: [0.0573, 0.1006, 0, 0.0573, 0]

Film B: [0, 0, 0, 0.0766, 0]

**Hitung Dot Product:**

$$\begin{aligned} A \cdot B &= (0,0573 \times 0) + (0,1006 \times 0) + (0 \times 0) + (0,0573 \times 0,0766) \\ &\quad + (0,0573 \times 0) + (0 \times 0) = 0,00438918 \end{aligned}$$

**Hitung Magnitude:**

$$\|A\| = \sqrt{(0,0573^2) + (0,1006^2) + (0,0573^2) + (0,0573^2)} = 0,1413$$

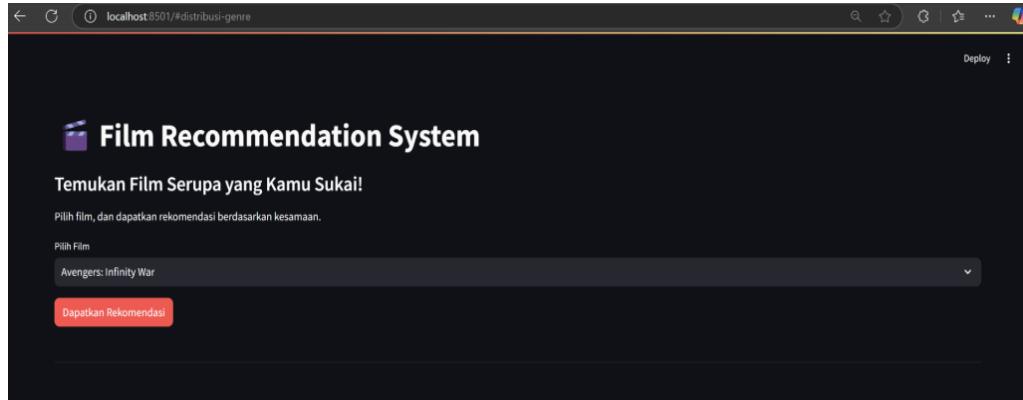
$$\|B\| = \sqrt{(0,0766^2)} = 0,0766$$

**Hitung Cosine Similarity:**

$$\text{Cosine Similarity}(A,B) = \frac{0,00438918}{0,1413 \times 0,0766} = 0,405$$

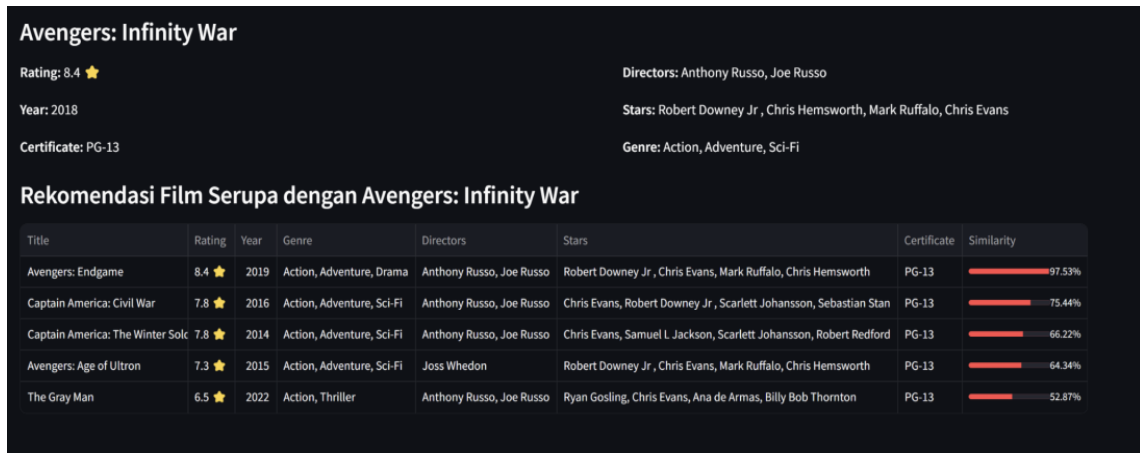
Jadi nilai kesamaan antara vektor A dan B adalah 0,4055 atau sekitar 40,55%.

## Implementasi



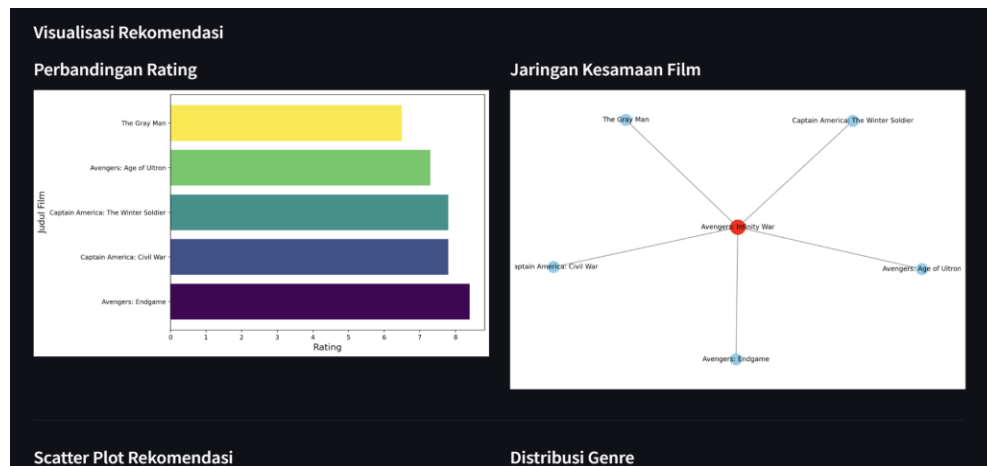
GAMBAR 3. Implementasi Halaman Web untuk Rekomendasi Film

Pada Gambar 3 merupakan tampilan web untuk dapat digunakan pencarian rekomendasi. Dicontohkan judul film yang dipilih adalah “Avengers Infinity War” sehingga akan memperoleh rekomendasi film yang mirip dengan judul film yang dipilih. Hasil tampilan film yang direkomendasikan dapat dilihat pada Gambar 4.



GAMBAR 4. Hasil Rekomendasi Film

Terlihat pada Gambar 4 bahwa 5 judul film yang direkomendasikan dari judul yang dipilih yaitu “Avengers: Infinity War” dengan nilai kemiripan tertinggi adalah 97,53% yaitu dengan judul “Avengers: Endgame”



GAMBAR 5. Visualisasi Hasil Rekomendasi

Pada Gambar 5, ditampilkan visualisasi rekomendasi film berdasarkan pilihan pengguna. Pertama, terdapat perbandingan rating antara film-film rekomendasi, yang membantu pengguna melihat perbedaan kualitas berdasarkan penilaian. Kedua, ditampilkan jaringan kesamaan film yang memperlihatkan hubungan atau keterkaitan antara film yang dipilih dengan film-film rekomendasi dalam bentuk graf atau jaringan. Ketiga, terdapat scatter plot rekomendasi yang menampilkan distribusi film-film rekomendasi berdasarkan dua parameter tertentu, seperti rating dan tahun rilis, untuk memberikan gambaran visual tentang pola atau kelompok film. Terakhir, ditampilkan distribusi genre dari film-film rekomendasi, yang memperlihatkan proporsi genre yang dominan atau variasi genre dalam rekomendasi.

## KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa sistem rekomendasi film berbasis Content-Based Filtering dengan menggunakan metode Cosine Similarity merupakan solusi yang efektif untuk memperoleh rekomendasi film. Sistem rekomendasi ini bekerja dengan menganalisis karakteristik atau fitur film, seperti genre, sutradara, aktor, dan tema cerita. Sistem rekomendasi pada penelitian ini menggunakan metode TF-IDF untuk memberikan bobot pada fitur-fitur tersebut serta Cosine Similarity digunakan untuk mengukur kesamaan antara film yang dipilih pengguna dengan film lainnya berdasarkan dataset. Hasilnya, sistem dapat memberikan rekomendasi film yang relevan dan personal sesuai dengan preferensi pengguna.

## DAFTAR PUSTAKA

1. S. Lestari and M. M. Ramdhani, "Sistem Rekomendasi film menggunakan metode content-based filtering studi KASUS materi data mining di SMK IDN boarding school," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 1581–1587, Sep. 2023. doi:10.35870/jimik.v4i3.381
2. M. I. Wardah and S. D. Putra, "Implementasi Machine Learning Untuk Rekomendasi Film Di Imdb Menggunakan Collaborative Filtering Berdasarkan Analisa Sentimen IMDB," *Jurnal Manajemen Informatika Jayakarta*, vol. 2, no. 3, p. 243, Jul. 2022, doi: <https://doi.org/10.52362/jmijayakarta.v2i3.868>.
3. Ananto Dwi Saputro and F. Amin, "Sistem Rekomendasi Content-Based Filtering Skincare Pria Di E-Commerce Shopee," *INTECOMS Journal of Information Technology and Computer Science*, vol. 7, no. 1, pp. 106–113, Jan. 2024, doi: <https://doi.org/10.31539/intecom.v7i1.8036>.
4. D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmaddeni, and D. Wulandari, "Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix: Implementation of Decision Tree Algorithm for Movie Recommendation and Rating Classification on the Netflix Platform," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 570–579, Mar. 2024, doi: <https://doi.org/10.57152/malcom.v4i2.1255>.



5. V. Sandrya, W. Wasino, and D. Arisandi, "Sistem Rekomendasi Film Menggunakan Metode Multiple Attribute Utility Theory," *Computatio : Journal of Computer Science and Information Systems*, vol. 6, no. 1, pp. 19–30, Jun. 2022, doi: <https://doi.org/10.24912/computatio.v6i1.17081>.
6. Muhammad Aryuska Pradana and A. T. Wibowo, "Movie Recommendation System Using Hybrid Filtering With Word2vec And Restricted Boltzmann Machines," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 1, pp. 231–241, Feb. 2024, doi: <https://doi.org/10.29100/jupi.v9i1.4306>.
7. A. Nilla and E. B. Setiawan, "Film Recommendation System Using Content-Based Filtering and the Convolutional Neural Network (CNN) Classification Methods," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 10, no. 1, p. 17, Feb. 2024, doi: <https://doi.org/10.26555/jiteki.v9i4.28113>.
8. A. R. Fitrianti, A. Rohmani, and W. Widjanarto, "Sistem Rekomendasi Film Berbasis Website Dengan Metode Prototype Menggunakan Algoritma K-Nearest Neighbors (KNN)," *JOINS (Journal of Information System)*, vol. 5, no. 2, pp. 278–287, Nov. 2020, doi: <https://doi.org/10.33633/joins.v5i2.4168>.
9. M. Yunus, "TF-IDF (Term Frequency-Inverse Document Frequency) : Representasi Vector Data Text," *Medium*, May 20, 2020. <https://yunusmuhammad007.medium.com/tf-idf-term-frequency-inverse-document-frequency-representasi-vector-data-text-2a4eff56cda>
10. E. Salim, J. Pragantha, and M. Lauro, "Perancangan Sistem Rekomendasi Film menggunakan metode Content- based Filtering." Available: [https://linter.untar.ac.id/repository/penelitian/buktipenelitian\\_10390001\\_7A281222103549.pdf](https://linter.untar.ac.id/repository/penelitian/buktipenelitian_10390001_7A281222103549.pdf)