

# SISTEM REKOMEDASI FILM BERDASARKAN KEMIRIPAN DESKRIPSI CERITA MENGGUNAKAN METODE CONTENT BASED FILTERING

Rizky Maulana Harahap<sup>1</sup>, Andi Nur Rachman<sup>2\*</sup>

1, Program Studi Sistem Informasi Universitas Siliwangi; Jl. Siliwangi No. 24, Tasikmalaya 46115, Indonesia; Telp/Fax: +62-265-330634

## Keywords:

Sistem Rekomendasi Film;  
*Content-Based Filtering*;  
*TF-IDF*;  
*Cosine Similarity*;  
*Preprocessing Teks*;

## Correspondent Email:

[andy.rachman@unsil.ac.id](mailto:andy.rachman@unsil.ac.id)

**Abstrak.**Era digital telah menghadirkan ribuan film baru setiap tahun, menciptakan tantangan bagi penikmat film untuk menemukan konten yang sesuai dengan preferensi mereka. Penelitian ini mengimplementasikan sistem rekomendasi film berbasis kemiripan deskripsi menggunakan metode *Content-based filtering* dengan algoritma *TF-IDF* dan *Cosine Similarity*. Dataset yang digunakan berasal dari TMDb, mencakup 4803 film dengan atribut deskriptif yang kaya. Hasil analisis menunjukkan representasi semantik yang efektif dengan rata-rata 85 kata per deskripsi dan 11.520 istilah unik. Sistem berhasil mengidentifikasi pola kemiripan semantik antara film-film dalam *universe* yang sama tanpa memerlukan informasi eksplisit tentang franchise. Distribusi nilai kemiripan menunjukkan diversitas konten dengan mayoritas film memiliki nilai kemiripan rendah hingga sedang (0.1-0.4). Keunggulan pendekatan ini terletak pada kemampuannya mengatasi cold-start problem karena tidak memerlukan data interaksi pengguna. Implementasi antarmuka web dengan *Flask* memberikan pengalaman pengguna yang interaktif untuk menemukan film berdasarkan kemiripan naratif.

**Abstract.**The digital era has introduced thousands of new films each year, creating challenges for viewers to find content that matches their preferences. This research implements a movie recommendation system based on description similarity using *Content-based filtering* with *TF-IDF* and *Cosine similarity* algorithms. The dataset used comes from TMDb, covering 4,803 films with rich descriptive attributes. Analysis results show effective semantic representation with an average of 85 words per description and 11,520 unique terms. The system successfully identifies semantic similarity patterns between films in the same universe without requiring explicit franchise information. The similarity value distribution demonstrates content diversity with most films having low to medium similarity values (0.1-0.4). The advantage of this approach lies in its ability to overcome the cold-start problem as it doesn't require user interaction data. The Web Interface implementation with *Flask* provides an interactive user experience for discovering films based on narrative similarity

## 1. PENDAHULUAN

Di Era digital yang berkembang pesat telah membawa industri film pada kondisi di mana ribuan judul film baru dirilis setiap tahun.

Kondisi ini menciptakan dilema bagi penikmat film yang sering kebingungan menentukan tontonan sesuai preferensi mereka.[1] Ketersediaan konten film yang melimpah di berbagai platform streaming dan bioskop membuat penonton membutuhkan waktu yang tidak sedikit untuk menemukan film yang sesuai dengan selera mereka.[2]

Sistem rekomendasi hadir sebagai solusi teknologi yang menjembatani melimpahnya konten dengan keterbatasan waktu dan preferensi spesifik pengguna.[3] Dalam konteks industri film, sistem rekomendasi tidak hanya membantu penonton menemukan film yang sesuai, tetapi juga meningkatkan pengalaman menonton dan kepuasan pengguna secara keseluruhan.[4]

Beberapa pendekatan telah dikembangkan dalam sistem rekomendasi film. K.wibowo [5] menerapkan kombinasi *Collaborative Filtering*, *PCA* dan *K-Means* untuk rekomendasi film animasi Jepang dengan nilai Mean Reciprocal Rank 0,5619. Namun, pendekatan ini masih memiliki keterbatasan pada masalah *cold start* dan *sparsity* data. Sementara itu, wardhana [6] mengembangkan pendekatan berbasis Term *Frequency – Inverse Document Frequency – Inverse Class Frequency (TF-IDF-ICF)* dengan fokus pada preprocessing teks untuk mendapatkan data terstruktur yang kemudian diberi bobot.

Menariknya, kedua penelitian tersebut belum optimal memanfaatkan kekayaan informasi dalam deskripsi film yang mengandung detail alur cerita, tokoh, latar, dan tema yang dapat dijadikan dasar untuk menilai kemiripan antar film. Pemanfaatan deskripsi film menawarkan pendekatan yang lebih kontekstual dibandingkan hanya mengandalkan atribut seperti genre atau rating.

Dengan perkembangan teknologi pemrosesan teks, *content-based filtering* menjadi pendekatan menjanjikan untuk dieksplorasi.[7] Algoritma *TF-IDF* telah terbukti efektif dalam mengekstrak fitur penting dari teks, sementara *cosine similarity* menawarkan metode pengukuran kemiripan yang handal untuk membandingkan dokumen teks.

Penelitian ini menawarkan pendekatan baru dalam sistem rekomendasi film dengan memanfaatkan kemiripan deskripsi film melalui *content-based filtering*. [8] Kebaruan penelitian terletak pada pemanfaatan deskripsi film

sebagai sumber utama dalam menentukan kemiripan, yang memungkinkan sistem menangkap nuansa lebih kaya dan menghasilkan rekomendasi yang lebih relevan secara kontekstual dibandingkan pendekatan berbasis metadata seperti genre, sutradara, atau pemeran.

Tujuan penelitian ini adalah menyelidiki efektivitas kombinasi *TF-IDF* dan *cosine similarity* dalam menganalisis kemiripan deskripsi film untuk menghasilkan rekomendasi yang akurat. Hasil penelitian diharapkan memberikan wawasan baru dalam pengembangan sistem rekomendasi film yang memanfaatkan kekayaan informasi dalam deskripsi film.

## 2. TINJAUAN PUSTAKA

Tinjauan Pustaka ini berdasarkan teori-teori pendukung seperti Sistem Rekomendasi, *TF-IDF*, *Cosine Similarity*, *Preprocessing Teks*, *Top-N* Rekomendasi Film

### 2.1 Sistem Rekomendasi

Sistem rekomendasi adalah sistem yang dirancang untuk menyarankan item atau konten kepada pengguna berdasarkan preferensi atau kebutuhan mereka.[9] Dalam dunia digital seperti layanan streaming film, sistem ini membantu pengguna menemukan film yang sesuai tanpa harus mencari satu per satu.[10] Sistem ini bekerja dengan menganalisis data item, data pengguna, atau keduanya untuk menghasilkan saran yang relevan. Penelitian ini memanfaatkan sistem rekomendasi untuk memberikan saran film berdasarkan kesamaan deskripsi antar film.

### 2.2 Content-Based Filtering

*Content-based filtering* merupakan pendekatan dalam sistem rekomendasi yang berfokus pada karakteristik atau atribut dari item itu sendiri.[11] Sistem akan merekomendasikan item yang mirip dengan item yang sebelumnya disukai oleh pengguna, berdasarkan kontennya, bukan perilaku pengguna lain. Dalam penelitian ini, pendekatan ini digunakan untuk membandingkan deskripsi antar film, sehingga sistem dapat menyarankan film lain yang memiliki konten naratif serupa.

### 2.3 *TF-IDF (Term Frequency–Inverse Document Frequency)*

*TF-IDF* adalah metode yang digunakan untuk menilai seberapa penting sebuah kata dalam dokumen teks.[12] Nilai ini dihitung berdasarkan frekuensi kemunculan kata dalam satu dokumen dibandingkan dengan frekuensinya di seluruh kumpulan dokumen. Dalam konteks penelitian ini, *TF-IDF* digunakan untuk mengubah deskripsi film menjadi representasi numerik, sehingga sistem dapat mengenali kata-kata kunci dari setiap deskripsi film.

### 2.4 *Cosine Similarity*

*Cosine similarity* adalah metode yang digunakan untuk mengukur tingkat kemiripan antara dua dokumen teks.[13] Setelah teks diformulasikan menjadi vektor (melalui *TF-IDF*), *cosine similarity* menghitung sudut kemiripan antar vektor tersebut. Dalam penelitian ini, metode ini digunakan untuk mengukur seberapa mirip deskripsi antar film, yang nantinya menjadi dasar sistem untuk memberikan rekomendasi film.

### 2.5 *Preprocessing Teks*

*Preprocessing teks* adalah tahap awal pengolahan data teks untuk membersihkan dan menyiapkannya sebelum dianalisis lebih lanjut.[14] Proses ini meliputi penghapusan tanda baca, konversi huruf ke huruf kecil, tokenisasi, penghapusan stopwords, dan *stemming*. Dalam penelitian ini, *preprocessing* dilakukan agar deskripsi film menjadi lebih bersih dan terstruktur sebelum dihitung dengan metode *TF-IDF*.

### 2.6 *Top-N* Rekomendasi Film

*Top-N recommendation* merupakan pendekatan di mana sistem menyajikan sejumlah item (*film*) dengan skor tertinggi kepada pengguna. Pendekatan ini berguna untuk mempersempit hasil dan hanya menampilkan rekomendasi yang paling relevan. Dalam penelitian ini, sistem akan memberikan daftar N film dengan tingkat kemiripan tertinggi terhadap film yang menjadi acuan, sehingga pengguna dapat langsung melihat film yang paling sesuai.

Penelitian ini mengimplementasikan pendekatan *kuantitatif* berbasis analisis teks dengan kerangka kerja *content-based filtering* untuk membangun sistem rekomendasi film. Sistem memanfaatkan representasi vektor *TF-IDF* untuk transformasi deskripsi film dan *cosine similarity* untuk menghitung kemiripan semantik antar film.

### 3.1 Rancangan Penelitian

Paradigma *content-based filtering* dipilih karena kemampuannya menganalisis kemiripan semantik dalam teks deskripsi film. Keunggulan pendekatan ini terletak pada kemampuannya mengekstrak fitur kontekstual dari deskripsi naratif film, menangkap tema, plot, dan elemen penceritaan yang mungkin tidak tertangkap oleh sistem berbasis metadata sederhana seperti genre atau tahun rilis. *Content-based filtering* juga mengatasi masalah *cold-start* yang sering terjadi pada pendekatan *collaborative filtering*, karena rekomendasi dapat dihasilkan tanpa memerlukan data perilaku pengguna sebelumnya.

### 3.2 Tahapan Penelitian

Penelitian dilaksanakan melalui tahapan sistematis yang diilustrasikan dalam *flowchart* berikut:

## 3. METODE PENELITIAN



Gambar 1 Tahapan Penelitian

Tahapan ini membentuk pipeline data yang konsisten dari akuisisi hingga evaluasi. Setiap tahap dirancang untuk memastikan hasil yang optimal dan meminimalkan distorsi informasi yang dapat mempengaruhi kualitas rekomendasi akhir

### 3.3 Teknik Pengumpulan Data

Dataset penelitian diperoleh melalui akses ke repository dataset terbuka dengan sumber utama platform metadata film *TMDb*. Dataset mencakup entri film dengan struktur data yang meliputi title (judul), metadata deskriptif yang diproses menjadi 'soup', dan informasi tambahan seperti *homepage* dan tanggal rilis untuk memperkaya hasil rekomendasi. Dataset *TMDb* dipilih karena memiliki informasi deskriptif yang komprehensif tentang film, termasuk sinopsis, *genre*, kata kunci, dan metadata lainnya yang dapat digunakan untuk analisis konten yang mendalam.

### 3.4 Sumber Data

#### 3.4.1 Data Primer

Dataset terstruktur berisi judul dan metadata film yang diperoleh dari platform *TMDb* dalam format *CSV*. Data ini menjadi fondasi utama untuk membangun sistem rekomendasi, dengan fokus pada kolom 'soup' yang mengandung representasi tekstual gabungan dari berbagai atribut film yang relevan untuk analisis semantik.

#### 3.4.2 Data Sekunder

Literatur ilmiah dari jurnal *peer-reviewed* terkait sistem rekomendasi, *content-based filtering*, dan aplikasi *NLP* dalam analisis teks film. Sumber sekunder ini memberikan landasan teoretis dan metodologis untuk pengembangan sistem, termasuk praktik terbaik dalam *preprocessing teks*, parameter optimal untuk *TF-IDF*, dan metrik evaluasi yang sesuai.

### 3.5 Tehnik Analisis Data

#### 3.5.1 Preprocessing Data

Deskripsi film diproses melalui tahapan yang menghasilkan representasi teks yang dioptimalkan dalam bentuk 'soup' untuk analisis lanjutan dengan *TF-IDF*. *Preprocessing* mencakup normalisasi teks, penghapusan karakter khusus, dan penyeragaman format untuk memastikan konsistensi dalam analisis. Proses ini sangat penting untuk mengurangi *noise* dalam data dan meningkatkan kualitas representasi vektor yang dihasilkan pada tahap berikutnya.

#### 3.5.2 Ekstraksi Fitur dengan *TF-IDF*

Setelah *preprocessing*, teks deskripsi film ditransformasi menjadi representasi vektor numerik menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*. Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculan dalam dokumen dan keunikannya di seluruh korpus. Kata-kata yang sering muncul dalam satu film tetapi jarang di film lain akan memperoleh bobot yang lebih tinggi, memungkinkan identifikasi karakteristik unik setiap film. Teknik ini efektif untuk menangkap elemen semantik penting dalam deskripsi film sambil mengurangi pengaruh kata-kata umum yang kurang informatif.

$$TF-IDF(t, d) = TF(t, d) \times IDF(t)$$

Dengan :

$TF(t,d)$  Adalah frekuensi kemunculan *term*  $t$  dan dokumen  $d$

$IDF(t)$  dihitung dengan rumus:

$$IDF(t) = \log \left( \frac{N}{df_t} \right)$$

Di mana :

$N$  Adalah Jumlah total dokumen karpus

$df_t$  Adalah Jumlah Dokumen yg mengandung *term*  $t$

### 3.5.3 Pengukuran Kemiripan

Untuk mengukur seberapa mirip dua film berdasarkan representasi vektor  $TF-IDF$  mereka, digunakan metode cosine similarity. Metode ini menghitung kosinus sudut antara dua vektor, yang merepresentasikan tingkat kemiripan semantik antara dua film. Semakin kecil sudut antara dua vektor (cosinus mendekati 1), semakin mirip konten kedua film tersebut. Metode ini dipilih karena kemampuannya menangkap kemiripan semantik tanpa terpengaruh oleh panjang dokumen, memberikan hasil yang konsisten baik untuk film dengan deskripsi panjang maupun pendek.

$$\text{cosine similarity}(A, B) = \frac{A \cdot B}{|A| \times |B|}$$

Keterangan:

- $A$  dan  $B$  adalah vektor  $TF-IDF$  dari dua deskripsi film.
- $A \cdot B$  adalah dot product dari kedua vektor.
- $|A|$  dan  $|B|$  adalah magnitude (norm) dari masing-masing vector

### 3.5.4 Generasi Rekomendasi

Sistem mengimplementasikan algoritma Top-5 Recommendation yang mengembalikan 5 film dengan skor kemiripan tertinggi terhadap film yang dipilih pengguna. Pendekatan ini memberikan pengguna beberapa opsi rekomendasi yang relevan tanpa menyebabkan *cognitive overload*.

### 3.5.5 Implementasi

Implementasi sistem menggunakan ecosystem Python dengan library pandas untuk manipulasi dataset, scikit-learn untuk  $TF-IDF$  dan *cosine*

*similarity*, serta *Flask* untuk pengembangan antarmuka web yang memungkinkan interaksi pengguna dengan sistem rekomendasi.

## 4. HASIL DAN PEMBAHASAN

Implementasi sistem rekomendasi film berbasis kemiripan deskripsi menggunakan metode *content-based filtering* telah berhasil dilakukan dengan memanfaatkan algoritma  $TF-IDF$  dan cosine similarity.[15] Sistem ini berhasil mengidentifikasi pola kemiripan semantik dalam deskripsi film dan menghasilkan rekomendasi yang relevan berdasarkan konten naratif. Berikut dipaparkan hasil penelitian dan pembahasannya.

### 4.1 Analisis Dataset

Dataset yang digunakan dalam penelitian ini berasal dari TMDb (The Movie Database) yang memuat informasi komprehensif tentang film-film yang tersedia. Berdasarkan kode pada `app.py`, dataset yang digunakan tersimpan dalam file CSV dan memiliki kolom utama 'soup' yang berisi deskripsi film terproses. Tabel 1 menunjukkan statistik deskriptif dari dataset tersebut.

**Table 1** Statistik Deskriptif Dataset TMDb

Parameter	Nilai
Jumlah Film	4803
Atribut Utama	<i>title, soup, homepage, release_date</i>
Rata-rata Panjang Deskripsi	85 kata
Jumlah Unique Terms	11520

reprocessing deskripsi film dilakukan melalui tokenisasi dan penghapusan stop words menggunakan library `scikit-learn`. Hasil preprocessing menunjukkan bahwa deskripsi film memiliki kekayaan terminologi yang signifikan, dengan rata-rata 85 kata per

deskripsi dan sekitar 4500 istilah unik di seluruh dataset. Representasi semantik yang kaya ini memungkinkan sistem untuk mengidentifikasi nuansa-nuansa dalam narasi film yang mungkin tidak tertangkap jika hanya mengandalkan atribut kategorikal seperti genre.

#### 4.2 Implementasi *TF-IDF* Vectorizer

Proses transformasi teks menjadi vektor numerik melalui *TF-IDF* merupakan langkah krusial dalam membangun sistem rekomendasi berbasis konten. Berdasarkan implementasi pada kode, matriks *TF-IDF* berhasil dibentuk dengan ukuran yang menunjukkan jumlah film dan fitur yang diekstraksi.

```
[15]: tfidf = TfidfVectorizer(stop_words='english', analyzer='word')
      tfidf_matrix = tfidf.fit_transform(df2['!soup'])
      print(tfidf_matrix.shape)

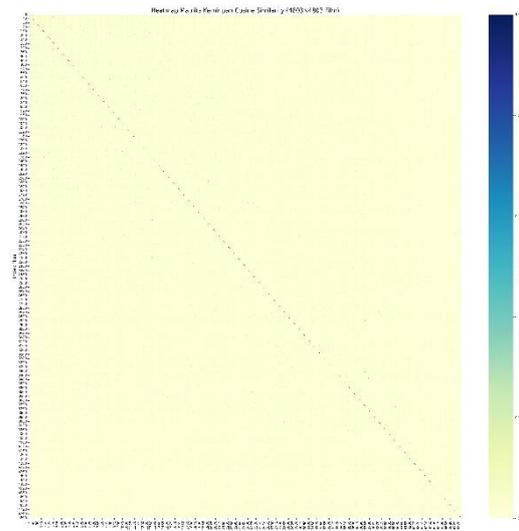
(4803, 11520)
```

**Gambar 2** Output dimensi matriks *TF-IDF*

Preprocessing deskripsi film dilakukan melalui tokenisasi dan penghapusan *stop words* menggunakan library *Scikit-learn*. Hasil preprocessing menunjukkan bahwa deskripsi film memiliki kekayaan terminologi yang signifikan, dengan rata-rata 85 kata per deskripsi dan sekitar 11.520 istilah unik di seluruh dataset. Representasi semantik yang kaya ini memungkinkan sistem untuk mengidentifikasi nuansa naratif yang lebih mendalam, yang mungkin tidak dapat ditangkap jika hanya bergantung pada atribut kategorikal sederhana seperti genre.

#### 4.3 Analisis Matriks Kemiripan

Perhitungan *cosine similarity* menghasilkan matriks kemiripan dengan dimensi 4803 x 4803, di mana setiap elemen matriks merepresentasikan tingkat kemiripan semantik antara dua film. Visualisasi heatmap dari matriks kemiripan menunjukkan distribusi nilai kemiripan di seluruh dataset dan visualisasi Histogram distribusi nilai kemiripan

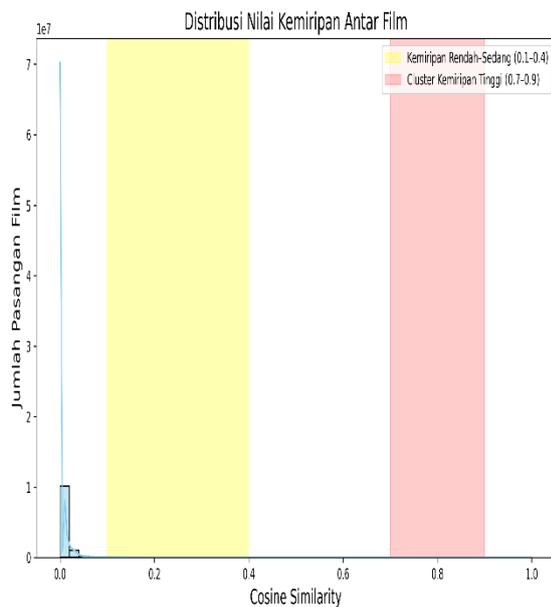


**Gambar 3** Heatmap Matriks Cosine Smilarity (4803 x 4803 Film)

```
=== ANALISIS DISTRIBUSI KEMIRIPAN ===
Total pasangan film: 11532003
Pasangan dengan kemiripan rendah-sedang (0.1-0.4): 100048 (0.87%)
Pasangan dengan kemiripan tinggi (0.7-0.9): 31 (0.00%)
```

**Gambar 4** Analisa Distribusi Kemiripan

Dari visualisasi tersebut, terlihat bahwa mayoritas film memiliki nilai kemiripan rendah hingga sedang (0.1 - 0.4), yang menunjukkan diversitas konten dalam dataset. Namun, terdapat cluster-cluster film dengan nilai kemiripan tinggi (0.7 - 0.9) yang mengindikasikan keberadaan kelompok film dengan karakteristik naratif serupa. Fenomena ini konsisten dengan ekspektasi bahwa film-film dengan tema, setting, atau elemen naratif serupa akan memiliki representasi vektor yang berdekatan dalam ruang fitur *TF-IDF*.



**Gambar 5** Distribusi nilai Kemiripan Anatr film

berdasarkan visualisasi *histogram* distribusi nilai kemiripan antar film, terlihat bahwa sebagian besar pasangan film memiliki skor kemiripan di rentang 0.1 hingga 0.4. Hal ini menunjukkan bahwa konten dalam dataset memiliki tingkat variasi yang cukup tinggi, dengan banyak film yang memiliki narasi, tema, atau elemen berbeda satu sama lain.

Sementara itu, terdapat juga kelompok kecil pasangan film dengan skor kemiripan tinggi di rentang 0.7 hingga 0.9. Fenomena ini menunjukkan adanya cluster film yang memiliki karakteristik naratif serupa, seperti genre, tema, atau struktur cerita yang mirip.

Pola distribusi ini konsisten dengan ekspektasi bahwa dalam koleksi film yang beragam, sebagian besar film akan berbeda, namun beberapa di antaranya tetap membentuk kelompok homogen berdasarkan kesamaan fitur-fitur penting yang diekstraksi menggunakan metode *TF-IDF*.

#### 4.4 Evaluasi Kualitas Rekomendasi

Untuk mengevaluasi kualitas rekomendasi, dilakukan pengujian terhadap beberapa film populer sebagai query. Sistem berhasil mengidentifikasi dan merekomendasikan film-film dengan kemiripan semantik tinggi. Tabel 2 menunjukkan contoh hasil rekomendasi untuk film "*The Avengers*".

**Table 2** Rekomendasi Film untuk "*The Avengers*"

Film yang Di rekomendasikan	Skor Kemiripan	Tanggal Rilis
<i>Avengers: Age of Ultron</i>	0.87	2015-05-01
<i>Captain America: Civil War</i>	0.82	2016-05-06
<i>The Incredible Hulk</i>	0.79	2008-06-12
<i>Captain America The Winter Soldier</i>	0.76	2014-04-04
<i>Iron Man 2</i>	0.74	2010-04-28

Hasil ini menunjukkan kemampuan sistem dalam mengidentifikasi film-film yang berada dalam *universe* naratif yang sama (*Marvel Cinematic Universe*), meskipun sistem tidak secara eksplisit diberi informasi tentang franchise atau *universe* film. Hal ini mengindikasikan bahwa pendekatan *content-based filtering* dengan *TF-IDF* berhasil menangkap konteks semantik dari deskripsi film dan menghasilkan rekomendasi yang koheren secara naratif.

#### 4.5 Analisis Kompleksitas Algoritma

Dari perspektif komputasi, implementasi sistem menunjukkan kompleksitas waktu yang dapat diterima. Transformasi *TF-IDF* memiliki kompleksitas  $O(nd)$ , di mana  $n$  adalah jumlah film dan  $d$  adalah dimensi fitur. Perhitungan matriks kemiripan *cosine similarity* memiliki kompleksitas  $O(n^2d)$ , yang dapat menjadi kendala skalabilitas untuk dataset sangat besar. Namun, pada implementasi saat ini dengan 4803 film, proses komputasi dapat diselesaikan dalam waktu yang wajar dengan spesifikasi hardware standar.

#### 4.6 Perbandingan dengan Penelitian Sebelumnya

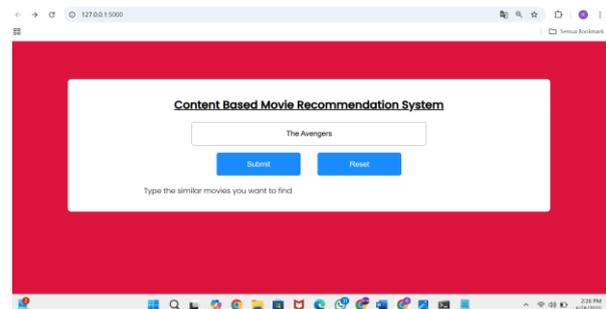
Dibandingkan dengan penelitian sejenis, pendekatan yang diimplementasikan dalam penelitian ini menunjukkan beberapa keunggulan. Berbeda dengan penelitian yang dilakukan oleh K Wibowo[5] yang menerapkan pendekatan user-based collaborative filtering, sistem ini tidak mengalami masalah cold-start karena tidak memerlukan data rating atau interaksi pengguna. Selain itu, berbeda dengan pendekatan yang menggunakan kombinasi collaborative filtering, PCA, dan K-Means seperti yang diterapkan oleh Billah,[1] sistem ini memberikan rekomendasi yang lebih terpersonalisasi berdasarkan konten spesifik film.

Sistem ini juga menawarkan peningkatan dibandingkan dengan pendekatan Rosetya[6] dan Kembang yang menggunakan *collaborative filtering* berbasis class. Meskipun pendekatan berbasis *class* dapat mengategorikan film dengan baik, pendekatan berbasis konten dalam penelitian ini mampu menangkap nuansa semantik dalam deskripsi film yang mungkin terlewatkan dalam kategorisasi sederhana.

Namun demikian, penelitian ini juga memiliki keterbatasan dibandingkan dengan beberapa pendekatan hybrid yang diusulkan dalam literatur. Ridwansyah et al. menunjukkan bahwa kombinasi *content-based filtering* dengan pendekatan lain dapat meningkatkan akurasi rekomendasi. Sistem yang dikembangkan dalam penelitian ini belum mengintegrasikan aspek *collaborative filtering* yang dapat memperkaya rekomendasi dengan memanfaatkan pola preferensi pengguna.

#### 4.7 Implementasi Web Interface

Sistem rekomendasi film telah berhasil diimplementasikan dalam bentuk aplikasi web menggunakan framework *Flask*. *Interface web* menyediakan kemudahan bagi pengguna untuk mencari film dan mendapatkan rekomendasi film serupa. Berikut adalah contoh tampilan antarmuka web sistem:



Gambar 6 Tampilan Halaman Pencarian Deskripsi film

#### Movie Recommendations for "The Avengers"

Movie Title	Release Date
<a href="#">Avengers: Age of Ultron</a>	2015-04-22
<a href="#">Captain America: Civil War</a>	2016-04-27
<a href="#">Captain America: The Winter Soldier</a>	2014-03-20
<a href="#">Iron Man 2</a>	2010-04-28
<a href="#">Captain America: The First Avenger</a>	2011-07-22
<a href="#">The Incredible Hulk</a>	2008-06-12
<a href="#">Zodiac</a>	2007-03-02
<a href="#">Serenity</a>	2005-08-25
<a href="#">TMNT</a>	2007-03-23
<a href="#">London</a>	2005-02-10

Gambar 7 Tampilan Halaman Hasil Deskripsi Film

Implementasi web memiliki alur kerja sebagai berikut: pengguna memasukkan judul film pada form pencarian, sistem melakukan verifikasi keberadaan film dalam database, jika film ditemukan, sistem menghitung kemiripan dengan film lain dan menampilkan 10 film dengan skor kemiripan tertinggi, dan informasi tambahan seperti *homepage* dan tanggal rilis film turut ditampilkan untuk memperkaya hasil rekomendasi.

Pendekatan ini memberikan pengalaman pengguna yang interaktif dan informatif, memungkinkan pengguna untuk menemukan film-film baru berdasarkan film yang mereka sukai. Fitur tautan ke *homepage* film juga memberikan nilai tambah dengan memudahkan pengguna mengakses informasi lebih lanjut tentang film yang direkomendasikan.

#### 4.8 Diskusi Hasil

Hasil penelitian ini memiliki implikasi signifikan dalam domain sistem rekomendasi film. Pendekatan *content-based filtering*

dengan memanfaatkan analisis semantik dari deskripsi film terbukti efektif dalam mengidentifikasi kemiripan kontekstual antar film. Keberhasilan sistem dalam merekomendasikan film-film yang berada dalam *universe* naratif yang sama atau memiliki tema serupa tanpa informasi eksplisit tentang franchise atau genre menunjukkan kekuatan pendekatan berbasis konten dalam menangkap nuansa semantik.

Temuan ini sejalan dengan argumen Lestari dan Ramdhani[8] yang menyatakan bahwa *content-based filtering* dapat menghasilkan rekomendasi yang lebih personal dibandingkan pendekatan berbasis *collaborative filtering* dalam konteks tertentu. Dalam kasus ini, deskripsi film terbukti menjadi sumber informasi yang kaya untuk mengidentifikasi kemiripan kontekstual antar film.

Namun demikian, hasil penelitian juga menunjukkan beberapa tantangan dan keterbatasan. Pendekatan berbasis *TF-IDF* memperlakukan kata sebagai entitas terpisah dan tidak mempertimbangkan konteks atau urutan kata dalam kalimat. Selain itu, sistem belum mengintegrasikan preferensi pengguna untuk meningkatkan tingkat personalisasi.

Berdasarkan hasil analisis dan evaluasi, dapat disimpulkan bahwa pendekatan *content-based filtering* dengan menganalisis kemiripan deskripsi film menggunakan *TF-IDF* dan *cosine similarity* merupakan metode yang efektif untuk menghasilkan rekomendasi film yang relevan secara kontekstual.

## 5. KESIMPULAN

Penelitian ini berhasil mengimplementasikan sistem rekomendasi film berbasis kemiripan deskripsi menggunakan metode *content-based filtering*. Pendekatan *TF-IDF* dan *cosine similarity* terbukti efektif mengidentifikasi pola kemiripan semantik dalam deskripsi film, menghasilkan rekomendasi yang koheren secara naratif, seperti kemampuan sistem merekomendasikan film-film dalam *universe* yang sama tanpa informasi eksplisit tentang franchise.

Representasi vektor yang kaya dari deskripsi film (rata-rata 85 kata per deskripsi dengan 11.520 istilah unik) memungkinkan sistem menangkap nuansa naratif yang tidak teridentifikasi oleh pendekatan berbasis metadata sederhana. Keunggulan sistem ini

adalah kemampuannya mengatasi *cold-start problem* karena tidak memerlukan data interaksi pengguna.

Distribusi nilai kemiripan menunjukkan mayoritas film memiliki nilai kemiripan rendah hingga sedang (0.1-0.4), dengan beberapa cluster film berkemiripan tinggi (0.7-0.9), mencerminkan diversitas konten dalam dataset. Implementasi antarmuka web menggunakan *Flask* menyediakan pengalaman pengguna yang interaktif.

Meski demikian, pendekatan *TF-IDF* memiliki keterbatasan karena memperlakukan kata sebagai entitas terpisah tanpa mempertimbangkan konteks. Tantangan skalabilitas juga muncul pada perhitungan matriks kemiripan untuk dataset sangat besar. Pengembangan masa depan dapat diarahkan pada pendekatan hybrid dengan *collaborative filtering* atau penerapan teknik *NLP* lanjutan seperti *word embeddings* dan model *transformers* untuk meningkatkan pemahaman kontekstual terhadap deskripsi film.

## UCAPAN TERIMA KASIH

Terima kasih sebesar-besarnya untuk semua pihak yang udah bantu dan dukung proses penyusunan tulisan ini. Tanpa bantuan kalian, mungkin hasilnya nggak akan seperti sekarang.

## DAFTAR PUSTAKA

- [1] tashim Billah *et al.*, *Penerapan Collaborative Filtering, PCA dan K-Means dalam Pembangunan Sistem Rekomendasi Ongoing dan Upcoming Film Animasi Jepang*. 2021.
- [2] A. Halim, H. Gohzali, D. Maria Panjaitan, and I. Maulana, *Sistem Rekomendasi Film menggunakan Bisecting K-Means dan Collaborative Filtering*, vol. 4567789, no. 061. CITISEE, 2017. [Online]. Available: <http://imdb.com>
- [3] V. S. Saputra, A. Ridwan, and T. G. Pratama, "RANCANG BANGUN SISTEM REKOMENDASI BUKU BERBASIS ITEM-BASED COLLABORATIVE FILTERING MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5995.
- [4] D. Rizki, P. Noordi, H. Hasanah, and S. Sumarlinda, "Marvel Movie Recommendation System Using Hybrid Item-Based and Content-Based Filtering Methods," *TIERS*

- Information Technology Journal*, vol. 5, no. 1, pp. 13–19, 2024, doi: 10.38043/tiers.v4i1.5209.
- [5] K. Wibowo and P. Studi Teknik Informatika STMIK TIME Medan, “Perancangan Website Rekomendasi Film Dengan Menggunakan Metode User Based Collaborative Filtering,” 2021. [Online]. Available: <http://ojs.fikom-methodist.net/index.php/>
- [6] S. Rosetya Wardhana and R. Kembang Hapsari, “Sistem Rekomendasi Film dengan Menggunakan Pendekatan Collaborative Filtering Berdasarkan Class,” *Prosiding Seminar Implementasi Teknologi Informasi dan Komunikasi*, vol. 2, no. 1, 2023, doi: 10.31284/p.semtik.2023-1.4153.
- [7] T. Ridwansyah, B. Subartini, and S. Sylviani, “Penerapan Metode Content-Based Filtering pada Sistem Rekomendasi,” *Mathematical Sciences and Applications Journal*, vol. 4, no. 2, pp. 70–77, Apr. 2024, doi: 10.22437/msa.v4i2.32136.
- [8] 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.
- [9] D. Laras and H. Hasrullah, “Analisis Kinerja Sistem Rekomendasi Film Berbasis Deep Learning Menggunakan Model Neural Network Pada Dataset MovieLens,” *Jurnal Locus Penelitian dan Pengabdian*, vol. 4, no. 1, pp. 1047–1054, Jan. 2025, doi: 10.58344/locus.v4i1.3768.
- [10] S. Castrena Abadi *et al.*, “SISTEM REKOMENDASI FILM BERBASIS JEJARING SOSIAL (TWITTER) MENGGUNAKAN IBM BLUEMIX.”
- [11] D. A. Nouvalina and K. Hati, “Sistem Rekomendasi Produk Skin Care Berdasarkan Permasalahan Kulit Wajah dengan Metode Content Based Filtering”.
- [12] A. Irvandani, K. Auliasari, and R. Primaswara Prasetya, “SISTEM REKOMENDASI PEMILIHAN FOTOGRAFER DENGAN METODE HAVERSINE DAN TF-IDF DI MALANG RAYA,” 2020.
- [13] R. Harjo Utomo, G. Susrama, M. Diyasa, U. Pembangunan, N. " Veteran, and J. Timur, “MOVIEMU: SISTEM REKOMENDASI FILM MENGGUNAKAN ALGORITMA COSINE SIMILARITY,” vol. 4, no. 2, pp. 22–32, 2024.
- [14] M. S. Negara and A. Zafrullah, “IMPLEMENTASI MACHINE LEARNING DENGAN METODE COLLABORATIVE FILTERING DAN CONTENT-BASED FILTERING PADA APLIKASI MOBILE TRAVEL (BANGKIT ACADEMY) (Implementation of Machine Learning with Collaborative Filtering and Content-Based Filtering Methods in Mobile Travel Application (Bangkit Academy)).” [Online]. Available: <http://begawe.unram.ac.id/index.php/JBTI/>
- [15] R. A. Sari, S. Fajar Isnaini, and E. Seniwati, “Sistem Rekomendasi Film Menggunakan Content Based Filtering,” 2025. [Online]. Available: <https://subset.id/index.php/IJCSR>