



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 1, No. 4, Tahun 2024

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 1, No. 4, Tahun 2024

Pages: 2135-2145

Sistem Rekomendasi Film Berbasis Konten Menggunakan Teknik
Cosine Similarity dan TF-IDF

Andra Husnul Azmi, Naufal Indra Mahardika, Rajawali Shaktika Anugrah
Prasetya, Anggraini Puspita Sari

Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur

Article in Journal of MISTER

Available at	: https://jurnal-serambimekkah.org/index.php/mister/index
DOI	: https://doi.org/10.32672/mister.v1i4.2206

How to Cite this Article

APA	: AZMI, A. H., MAHARDIKA, N. I., ANUGRAH PRASETYA, R. S., & PUSPITA SARI, A. (2024). Sistem Rekomendasi Film Berbasis Konten Menggunakan Teknik Cosine Similarity dan TF-IDF. <i>Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research</i> , 1(4), 2135 - 2145. https://doi.org/10.32672/mister.v1i4.2206
Others Visit	: https://jurnal-serambimekkah.org/index.php/mister/index

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



Sistem Rekomendasi Film Berbasis Konten Menggunakan Teknik Cosine Similarity dan TF-IDF

Andra Husnul Azmi¹, Naufal Indra Mahardika², Rajawali Shaktika Anugrah Prasetya³,
Anggraini Puspita Sari⁴

^{1,2,3,4}Informatika, Universitas Pembangunan Nasional "Veteran" Jawa Timur
Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur

*Email Korespondensi: anggraini.puspita.if@gmail.ac.id

Diterima: 01-09-2024

| Disetujui: 03-09-2024

| Diterbitkan: 05-09-2024

ABSTRACT

The goal of the research is to develop a content-based movie recommendation system that utilizes cosine similarity techniques to get better accuracy and relevance of recommendations to users. The methods used include text analysis of movie synopsis for feature extraction, representation vector generation, and cosine similarity calculation to determine the similarity between movies. Natural language processing is used to understand user preferences and compose recommendations accordingly. The results show that the developed recommendation system is able to enhance the accuracy of movie recommendations based on content, as well as provide a more personalized and relevant experience for users.

Key Words : Similarity, Movie, Content , Recommendation, Vector.

ABSTRAK

Tujuan untuk memahami preferensi pengguna dan menyusun rekomendasi yang sesuai. Penelitian ini menunjukkan hasil bahwa sistem rekomendasi yang dikembangkan mampu meningkatkan akurasi rekomendasi film berdasarkan konten, serta memberikan pengalaman yang lebih personal dan relevan. Dari penelitian ini adalah untuk mengembangkan sistem rekomendasi film berbasis konten yang memanfaatkan teknik cosine similarity untuk meningkatkan akurasi dan relevansi rekomendasi kepada pengguna. Metode yang digunakan meliputi analisis teks sinopsis film untuk ekstraksi fitur, pembuatan vektor representasi, dan perhitungan kesamaan kosinus untuk menentukan kemiripan antara film-film. Pemrosesan digunakan bagi pengguna.

Kata kunci: Similarity, Film, Konten, Rekomendasi, Vektor

PENDAHULUAN

Film adalah salah satu media komunikasi modern yang efektif dalam bentuk audio visual dan memiliki sifat yang sangat kompleks. Film adalah karya yang sangat unik dan menarik, karena menyampaikan ide dalam bentuk gambar bergerak dan berfungsi sebagai sumber informasi, hiburan, propaganda, serta politik. Selain itu, film juga dapat menjadi sarana rekreasi dan edukasi yang dapat dinikmati oleh Masyarakat[1]. Film telah menjadi salah satu bagian penting dalam kehidupan sehari-hari banyak orang. Dalam era modern yang dipenuhi dengan perkembangan teknologi, platform streaming untuk menonton film seperti netflix memiliki tingkat popularitas yang tinggi. Platform tersebut menyediakan kemudahan akses dan praktis untuk melihat film dari berbagai negara yang memungkinkan pengguna dapat menonton kapan saja dan dimana saja. Namun, dengan pertumbuhan teknologi yang pesat pada film, pengguna seringkali mendapatkan tantangan untuk mencari film baru yang sesuai dengan preferensi mereka [2].

Untuk mengatasi masalah pemilihan film, dibuatlah sistem rekomendasi film yang dapat membantu pengguna yang kebingungan dalam mencari film dengan memberikan rekomendasi yang berasal dari analisa data[3]. Sistem rekomendasi telah menjadi salah satu solusi penting untuk membantu pengguna mendapatkan film sesuai dari preferensi yang mereka suka. Metode yang biasanya dilakukan untuk menemukan rekomendasi film dengan cara mengandalkan riwayat data pengguna, seperti riwayat tontonan pengguna[4]. Terdapat banyak metode untuk melakukan pengolahan data untuk sistem rekomendasi film, tetapi penelitian ini berfokus pada dua metode, yaitu TF-IDF dan *Cosine Similarity*. TF-IDF (Frekuensi Istilah - Frekuensi Terbalik Dokumen) merupakan metode kuantitatif yang menilai seberapa penting suatu kata dalam sebuah dokumen dibandingkan dengan keseluruhan koleksi dokumen. Teknik ini berguna untuk menentukan kata kunci yang signifikan, yang dapat digunakan untuk identifikasi atau kategorisasi dokumen tertentu. *Cosine Similarity* merupakan ukuran yang lazim digunakan dalam pengambilan informasi dan area studi yang serupa. Ini menggambarkan dokumen teks sebagai vektor yang dibentuk dari istilah-istilah yang terkandung di dalamnya. Dengan menggunakan pendekatan ini, kesamaan antara dua dokumen dapat dievaluasi dengan menghitung sudut kosinus antara vektor istilah dari masing-masing dokumen[5].

Dengan menghitung kemiripan dari sinopsis suatu film, sistem akan memberikan rekomendasi film yang dianggap memiliki kesamaan antara satu film dengan film lainnya. Film yang tidak memiliki kesamaan dengan film yang dicari pengguna akan dihilangkan dalam proses pencarian. Hasil sistem yang diberikan berupa list film yang mempunyai kemiripan satu sama lain[6]

Penelitian ini melibatkan pengumpulan data dan pemrosesan teks terkait film dari sumber yang memiliki sinopsis suatu film. Metode ini diterapkan agar dapat mencari judul film yang memiliki kesamaan setelah dilakukannya perhitungan *Cosine Similarity*. Pengujian data akan dilakukan menggunakan dataset luas untuk menguji keefektifan sistem rekomendasi yang dibuat.

METODE PENELITIAN

Metode TF-IDF

TF-IDF (Term Frequency-Inverse Document Frequency) adalah metode dalam pemrosesan bahasa alami dan *text mining* yang digunakan untuk mengevaluasi pentingnya sebuah kata dalam sebuah dokumen relatif terhadap kumpulan dokumen lainnya. Metode ini menggabungkan dua ukuran utama: Term Frequency (TF) dan Inverse Document Frequency (IDF). Term Frequency mengukur seberapa sering

sebuah istilah muncul dalam sebuah dokumen, dengan rumus sederhana yaitu jumlah kemunculan istilah dibagi dengan total jumlah kata dalam dokumen tersebut. Sementara itu, Inverse Document Frequency mengukur seberapa penting sebuah istilah dengan mempertimbangkan jumlah dokumen yang mengandung istilah tersebut, di mana istilah yang umum muncul di banyak dokumen akan memiliki nilai IDF yang rendah. Frekuensi kemunculan kata dalam dokumen menunjukkan tingkat kepentingan kata tersebut dalam dokumen tersebut[7].

$$\text{IDF} = \log\left(\frac{D}{\text{DF}}\right)$$
$$\text{Weight} = \text{TF} \times \text{IDF}$$

Keterangan:
D = jumlah dokumen yang ada di dalam database
DF = banyak dokumen yang mengandung kata(*term*)
W = hasil bobot nilai dokumen *Frequency*
tf = banyaknya kata(*term*) pada suatu dokumen
idf = *Inversed Document Frequency*

Gambar 1. TF-IDF

Cosine Similarity

Cosine similarity adalah metode pemrosesan dalam bahasa alami dan *text mining* yang digunakan untuk mengukur kesamaan antara dua vektor teks, seperti dokumen atau kalimat, dengan menghitung kosinus dari sudut di antara mereka. Nilai *cosine similarity* berkisar antara -1 dan 1, di mana nilai 1 menunjukkan bahwa dua vektor teks tersebut identik, nilai 0 menunjukkan bahwa mereka tidak memiliki kesamaan, dan nilai -1 menunjukkan bahwa mereka berlawanan secara sempurna. Untuk menghitung tingkat kesamaan antar dokumen tetapi tidak menyertakan frekuensi kata (*term*). Metode *cosine similarity* tidak mempertimbangkan makna semantik dari setiap kata. Hal ini berbeda dengan distance based method, information content based method dan distance and information content based method. Dengan adanya pengukuran TF-IDF (Term Frequency – Inverse Document Frequency) yang dapat memberikan nilai ukur pada setiap kata dalam dokumen akan membantu *cosine similarity* untuk memproses kata secara maksimal [8][9].

Alur Penelitian

Alur penelitian yang dipakai dalam studi ini adalah sistem rekomendasi menggunakan *Cosine Similarity* dan *TF-IDF*. Dalam menyelesaikan kasus ini terdapat tahapan-tahapan yang akan dilaksanakan pada prosesnya, bisa terlihat pada gambar diagram alur dibawah



Gambar 2. Tahapan Metode Penelitian

A. Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang akan digunakan untuk membentuk sistem rekomendasi film. Pengumpulan data merupakan langkah awal yang sangat penting dalam pengembangan sistem rekomendasi film berbasis konten. Data tersebut berasal dari Kaggle yang telah menyediakan dataset film-film yang sudah ada[10]. Kaggle merupakan platform yang digunakan

untuk membuat model terbaik untuk menganalisa dan memprediksi suatu dataset. Dataset tersebut dapat diakses melalui web secara online.

Dataset yang digunakan dalam penelitian berasal dari Kaggle dengan judul “TMDB Movies Dataset” yang dapat diakses melalui web kaggle . Dataset berisi informasi tentang film dengan jumlah 10.000 data, termasuk judul, genre, dan sinopsis. Dataset yang telah diunduh kemudian dimuat ke dalam Python untuk diproses lebih lanjut. Berikut adalah beberapa contoh data yang terdapat pada dataset,

	id	title	genre	original_language	overview	popularity	release_date	vote_average	vote_count
0	278	The Shawshank Redemption	Drama,Crime	en	Framed in the 1940s for the double murder of h...	94.075	1994-09-23	8.7	21862
1	19404	Dilwale Dulhania Le Jayenge	Comedy,Drama,Romance	hi	Raj is a rich, carefree, happy-go-lucky second...	25.408	1995-10-19	8.7	3731
2	238	The Godfather	Drama,Crime	en	Spanning the years 1945 to 1955, a chronicle o...	90.585	1972-03-14	8.7	16280
3	424	Schindler's List	Drama,History,War	en	The true story of how businessman Oskar Schind...	44.761	1993-12-15	8.6	12959
4	240	The Godfather: Part II	Drama,Crime	en	In the continuing saga of the Corleone crime f...	57.749	1974-12-20	8.6	9811
5	667257	Impossible Things	Family,Drama	es	Matilde is a woman who, after the death of her...	14.358	2021-06-17	8.6	255

Gambar 3. Sampel Data

B. Pra-Pemrosesan

Pra-pemrosesan dalam penelitian ini bertujuan untuk mengolah data menjadi data siap pakai untuk proses selanjutnya[11]. Dalam penelitian ini, sinopsis film harus melalui serangkaian langkah pra-pemrosesan sebelum dimasukkan ke dalam model klasifikasi menggunakan TF-IDF. Tujuan dari pra-pemrosesan ini adalah untuk membersihkan teks, mengubah format, atau menyesuaikan data agar sesuai dengan kebutuhan model. Adapun langkah-langkah pra-pemrosesan yang dilakukan meliputi :

a. Pemilihan Kolom

Pemilihan kolom dilakukan agar data yang digunakan sesuai dengan yang diinginkan, kemudian data tersebut dapat berjalan secara efektif. Proses ini dilakukan menggunakan.

b. Penghapusan *Stopword*

Penghapusan *stopword* adalah salah satu upaya yang umum dilakukan dalam pra-pemrosesan data untuk menghilangkan kata-kata umum yang tidak digunakan dalam analisis teks[12]. Menghapus kata-kata ini dapat mengurangi kompleksitas data yang diolah, sehingga mempercepat analisis[13]. kata yang dihilangkan seperti “the”, “and”, “a” dalam Bahasa Inggris atau “di”, “dari”, “yang” dalam Bahasa Indonesia. Tujuannya adalah agar membersihkan teks dari kata-kata yang umum dan tidak relevan, sehingga menaikkan akurasi dalam keberlanjutan analisis.

C. Modeling

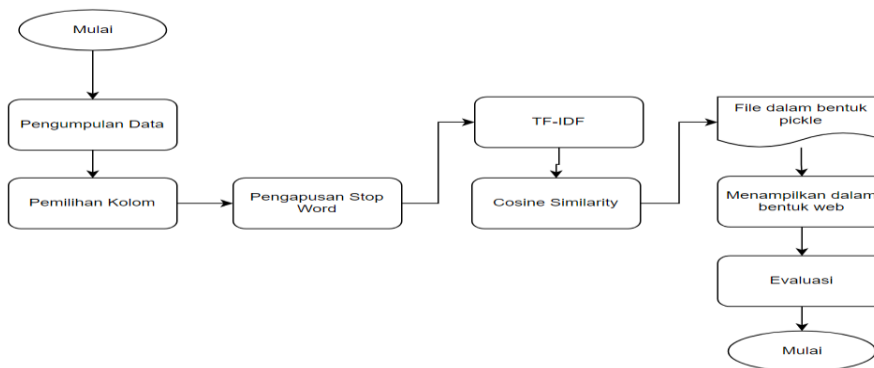
Modeling atau pemodelan adalah sketsa sistem nyata ke dalam suatu model yang mencerminkan tingkah laku, wujud fisik, dan karakteristik lain yang mirip dengan sistem nyata tersebut. Pemodelan adalah proses inti untuk pembuatan simulasi. Model yang baik adalah model yang efisien dalam merepresentasikan sistem yang ada[14]. Dalam konteks ini, pemodelan bertujuan untuk menganalisis kesamaan antara deskripsi film dalam dataset menggunakan metode *cosine similarity*. Teknik TF-IDF

digunakan untuk merepresentasikan teks deskripsi film sebagai vektor, dan *cosine similarity* dihitung untuk menghitung tingkat kesamaan antar vektor tersebut.

D. Evaluasi

Setelah melakukan serangkaian alur penelitian, dilakukan evaluasi. Evaluasi sendiri adalah proses yang terstruktur dan berkelanjutan untuk menilai kualitas hasil penelitian, menggunakan kriteria dan pertimbangan tertentu, guna menghasilkan keputusan yang informatif terkait dengan efektivitas, validitas, dan relevansi temuan penelitian tersebut[15].

HASIL DAN PEMBAHASAN



Gambar 4. Alur pembuatan sistem

Gambar 4 merupakan sebuah alur dalam membuat sistem rekomendasi film. Sebelum membuat sistem rekomendasi film, data yang sudah dikumpulkan diolah terlebih dahulu atau disebut Pra-Pemrosesan. Didalam Pra-Pemrosesan terdapat tiga proses yang harus dilewati, yaitu :

1. Pemilihan Kolom

Pemilihan kolom dilakukan agar memudahkan proses pengolahan data selanjutnya. Kolom yang akan diambil pada penelitian ini adalah kolom id, kolom title, dan kolom tags.

	id	title	tags
0	278	The Shawshank Redemption	Framed in the 1940s for the double murder of h...
1	19404	Dilwale Dulhania Le Jayenge	Raj is a rich, carefree, happy-go-lucky second...
2	238	The Godfather	Spanning the years 1945 to 1955, a chronicle o...
3	424	Schindler's List	The true story of how businessman Oskar Schind...
4	240	The Godfather: Part II	In the continuing saga of the Corleone crime f...

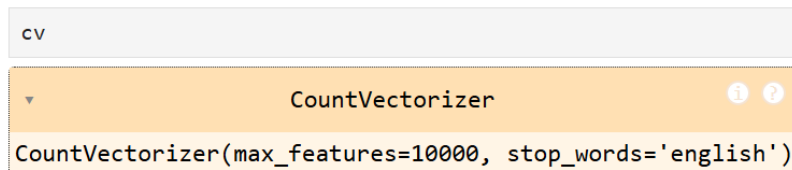
Gambar 5. Pemilihan Kolom

Pada gambar 5, ketiga kolom tersebut dipilih agar mempermudah perhitungan mengukur sistem rekomendasi. Kolom tags adalah kolom yang digunakan sebagai acuan melihat kesamaan judul film lainnya

2. Penghapusan *Stopword*

Penghapusan *stopword* dilakukan untuk menghapus semua kata yang tidak diperlukan dan proses pengolahan dan juga untuk mempermudah pengolahan data.

```
cv=CountVectorizer(max_features=10000, stop_words='english')
```



Gambar 6. Penghapusan *Stopword*

Pada gambar 6 bisa dilihat bahwa dilakukan proses penghapusan *stopword* untuk 10.000 data yang ada dalam dataset.

Setelah melakukan pra-pemrosesan data, langkah selanjutnya adalah menghitung TF-IDF dari tabel yang telah diproses sebelumnya. TF-IDF digunakan untuk menilai kata-kata mana yang penting untuk lanjutan dalam pengembangan sistem dan pengolahan data. Kata-kata yang dipilih kemudian akan diubah menjadi sebuah vektor yang berfungsi untuk mempersiapkan perhitungan kesamaan setelah proses ini.

```
array([[1.          , 0.05634362, 0.12888482, ..., 0.07559289, 0.11065667,
        0.06388766],
       [0.05634362, 1.          , 0.07624929, ..., 0.          , 0.03636965,
        0.          ],
       [0.12888482, 0.07624929, 1.          , ..., 0.02273314, 0.06655583,
        0.08645856],
       ...,
       [0.07559289, 0.          , 0.02273314, ..., 1.          , 0.03253   ,
        0.02817181],
       [0.11065667, 0.03636965, 0.06655583, ..., 0.03253   , 1.          ,
        0.0412393  ],
       [0.06388766, 0.          , 0.08645856, ..., 0.02817181, 0.0412393  ,
        1.          ]])
```

Gambar 7. TF-IDF

Gambar 7 menampilkan data yang sudah menjadi vektor setelah proses TF-IDF

Setelah melakukan proses TF-IDF, selanjutnya dilakukan proses *Cosine Similarity*. Proses ini bertujuan agar dapat melihat rekomendasi film berdasarkan kesamaan vektor yang didapatkan dari proses TF-IDF. Film yang memiliki kesamaan vektor dengan film yang dicari akan muncul dan memberitahu pengguna.

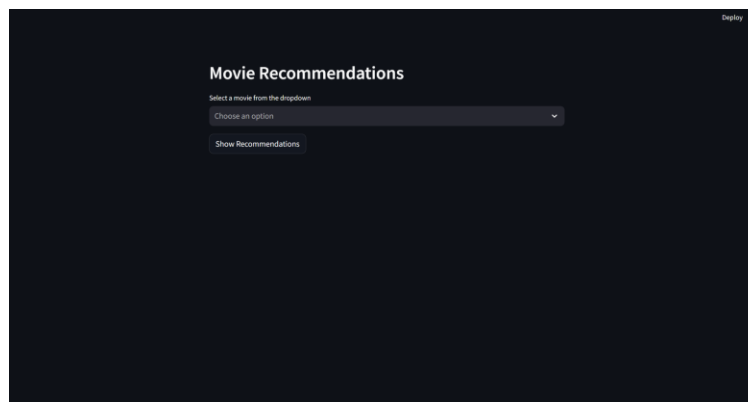
```
recommand("Batman")
```

```
Batman  
Batman Returns  
The Batman vs. Dracula  
Batman Beyond: Return of the Joker  
The Dark Knight
```

Gambar 8. Uji Rekomendasi Film

Pada gambar 8, telah diuji coba sebuah judul film yaitu "batman" untuk dicari film mana saja yang memiliki kemiripan dengan film tersebut. Hasilnya muncul lima list yang memiliki kesamaan dengan batman. Hasil lima list sesuai source code yang digunakan untuk mengeluarkan berapa banyak list yang digunakan. Kemudian, untuk hasil dari list tersebut berdasarkan urutan yang ada dalam tabel didalam dataset.

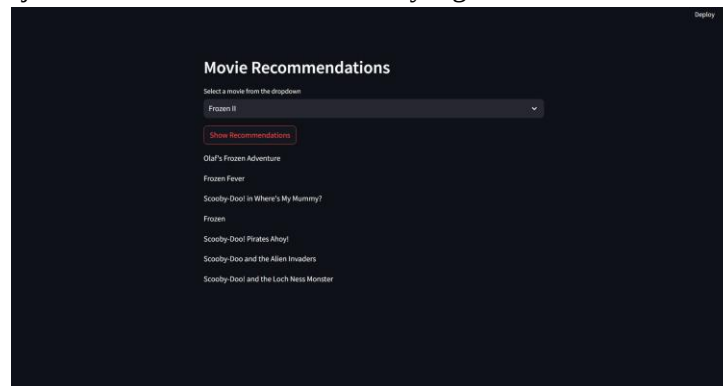
Setelah selesai melakukan pengujian untuk mencari film yang mirip dengan "Batman", data hasil pengolahan tersebut disimpan menggunakan format pickle. Format pickle digunakan untuk menyimpan objek Python, termasuk daftar film yang telah diproses dan direkomendasikan. Tujuan penyimpanan ini adalah untuk memungkinkan penggunaan data tersebut di dalam sebuah situs web yang dibangun menggunakan bahasa pemrograman Python. Situs web ini akan menampilkan hasil dari pengolahan data tersebut kepada pengguna.



Gambar 9. Tampilan Web Rekomendasi Film

Gambar 9 menampilkan sebuah situs web yang dibuat untuk memberikan pengalaman interaktif kepada pengguna dalam menerima rekomendasi film yang sesuai dengan preferensi mereka. Situs ini menggunakan teknologi Python, khususnya framework Streamlit, yang dirancang untuk menyajikan antarmuka web yang mudah digunakan dan responsif. Melalui platform ini, data film diproses dan ditampilkan secara dinamis, memungkinkan pengguna untuk menemukan rekomendasi film berdasarkan analisis data yang telah dilakukan sebelumnya.

Setelah selesai membuat tampilan sistem rekomendasi film, selanjutnya dilakukan evaluasi pada sistem, sistem akan diuji coba untuk mencari suatu film yang memiliki kesamaan dengan film lain.



Gambar 10. Evaluasi dan Uji Web Rekomendasi Film

Pada gambar 10, terlihat bahwa sistem melalui situs web berhasil menampilkan rekomendasi film. Di situs web tersebut, film-film direkomendasikan dalam bentuk sebuah daftar yang terdiri dari tujuh film. Film-film ini dipilih karena memiliki kesamaan vektor dengan film yang sedang dicari oleh pengguna. Ini menunjukkan bahwa situs web menggunakan teknologi yang efektif untuk menganalisis dan memproses data film, kemudian menyajikan hasilnya kepada pengguna dalam format yang mudah dipahami dan digunakan. Pendekatan ini memungkinkan pengguna untuk dengan cepat menemukan rekomendasi film yang sesuai dengan preferensi atau pencarian mereka.

KESIMPULAN

Penelitian ini fokus pada pengembangan sistem rekomendasi film yang mengimplementasikan metode *cosine similarity* dan TF-IDF. Metode ini digunakan untuk menganalisis dataset yang terdiri dari 10.000 film, yang telah melalui proses pra-pemrosesan untuk mempersiapkan data sebelum dilakukan perhitungan persamaan vektor. Pada tahap implementasi, *cosine similarity* digunakan untuk mengukur seberapa mirip vektor representasi film yang satu dengan yang lain berdasarkan pada fitur-fitur yang dihasilkan dari TF-IDF. Hasil dari perhitungan *cosine similarity* ini kemudian digunakan untuk memberikan rekomendasi kepada pengguna. Rekomendasi ini menunjukkan film-film yang memiliki kesamaan yang tinggi dengan film yang disukai atau dipilih oleh pengguna. Dengan menggunakan pendekatan ini, sistem rekomendasi dapat memberikan hasil yang lebih relevan dan personal untuk pengguna, memungkinkan mereka menemukan film-film baru yang memiliki kesamaan dalam karakteristik atau tema dengan film yang mereka sukai sebelumnya.

SARAN

Berdasarkan hasil penelitian yang telah dilakukan, berikut beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya serta pengembangan lebih lanjut dari sistem rekomendasi film:

1. **Peningkatan Volume Data:** Meskipun penelitian ini menggunakan dataset dengan 10.000 film, penggunaan dataset yang lebih besar dapat memberikan hasil yang lebih representatif dan

meningkatkan akurasi rekomendasi. Penambahan data dari berbagai sumber, termasuk berbagai platform streaming, dapat membantu memperluas jangkauan dan keragaman film yang direkomendasikan.

2. **Penggunaan Teknik Pembelajaran Mesin:** Selain *cosine* similarity dan TF-IDF, disarankan untuk mengeksplorasi teknik-teknik pembelajaran mesin yang lebih canggih seperti neural networks, deep learning, dan collaborative filtering. Teknik-teknik ini dapat menangkap pola yang lebih kompleks dan memberikan rekomendasi yang lebih personal dan relevan.
3. **Integrasi Data Eksternal:** Integrasi data dari sumber eksternal seperti ulasan pengguna, rating, dan popularitas di media sosial dapat memberikan konteks tambahan yang berharga untuk meningkatkan kualitas rekomendasi. Informasi ini dapat membantu sistem untuk memahami tren dan preferensi pengguna secara lebih mendalam.
4. **Pengembangan Antarmuka Pengguna:** Pengembangan antarmuka pengguna yang intuitif dan interaktif dapat meningkatkan pengalaman pengguna dalam menerima rekomendasi. Fitur-fitur seperti filter genre, tahun rilis, dan aktor favorit dapat memberikan kontrol lebih kepada pengguna untuk menyesuaikan rekomendasi sesuai dengan preferensi mereka.
5. **Evaluasi dan Validasi Berkelanjutan:** Melakukan evaluasi dan validasi berkelanjutan terhadap sistem rekomendasi sangat penting untuk memastikan keandalannya. Penggunaan matrik evaluasi seperti Precision, Recall, dan F1-Score dapat memberikan gambaran yang jelas tentang performa sistem dan area yang memerlukan perbaikan.
6. **Penerapan Aspek Semantik:** Mengintegrasikan analisis semantik untuk memahami konteks dan makna dari konten film dapat meningkatkan relevansi rekomendasi. Pendekatan seperti Natural Language Processing (NLP) dapat digunakan untuk menganalisis deskripsi film dan ulasan pengguna secara lebih mendalam.
7. **Pengujian dengan Berbagai Segmentasi Pengguna:** Menguji sistem rekomendasi dengan berbagai segmentasi pengguna, seperti berdasarkan umur, jenis kelamin, dan preferensi genre, dapat memberikan wawasan tentang bagaimana sistem bekerja pada berbagai kelompok pengguna. Hal ini dapat membantu dalam menyesuaikan algoritma agar lebih responsif terhadap kebutuhan spesifik tiap segmen.

Dengan mempertimbangkan saran-saran di atas, diharapkan penelitian lebih lanjut dapat memperbaiki dan menyempurnakan sistem rekomendasi film, sehingga mampu memberikan pengalaman menonton yang lebih baik dan lebih disesuaikan dengan preferensi individu pengguna.

Penelitian ini bukanlah produk akhir, dikarenakan teknologi terus berkembang seiring waktu. Diharapkan dengan adanya paper ini dapat membantu apabila ada pembuat yang ingin mengembangkan aplikasi yang mungkin serupa dengan sistem rekomendasi kami.

DAFTAR PUSTAKA

- [1] C. Nugraha, I. Fitri Astuti, and A. Harsa Kridalaksana, "Movie Organizer Menggunakan Teknik Web Scrapping," *J. Inform. Mulawarman*, vol. 9, no. 3, pp. 56–61, 2014.
- [2] N. R. Fitradhi, M. Firman Hidayat, T. W. Saputro, M. G. Alifian, and A. P. Sari, "LAPORAN FINAL PROJECT MACHINE LEARNING REKOMENDASI MUSIK SPOTIFY MENGGUNAKAN METODE K-MEANS," 2023.

- [3] Intan Hervianda Putri, Siti Muslimah K. H. Nurakhmadyavi, and Erwin Eko Wahyudi, "Literature Review: Sistem Rekomendasi untuk Buku dan Film," *Pros. Semin. Nas. Tek. Elektro, Sist. Informasi, dan Tek. Inform.*, vol. 1, no. 1, pp. 75–81, 2022, [Online]. Available: <https://ejurnal.itats.ac.id/snestikdanhttps://snestik.itats.ac.id>
- [4] M. I. Fikri, R. Amin, M. Rafly Amartharizqi, F. A. Sofi, and A. P. Sari, "Seminar Nasional Informatika Bela Negara (SANTIKA) Sistem Rekomendasi Musik Berdasarkan Preferensi Pengguna Dengan Menggunakan Metode Natural Language Processing (NLP)".
- [5] R. Al Rasyid and D. H. U. Ningsih, "Penerapan Algoritma TF-IDF dan Cosine Similarity untuk Query Pencarian Pada Dataset Destinasi Wisata," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 8, no. 1, pp. 170–178, 2024, doi: 10.35870/jtik.v8i1.1416.
- [6] M. Fajriansyah, P. P. Adikara, and A. W. Widodo, "Sistem Rekomendasi Film Menggunakan Content Based Filtering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 6, pp. 2188–2199, 2021, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/9163>
- [7] M. Yusuf and A. Cherid, "Implementasi Algoritma Cosine Similarity Dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar," *J. Ilm. Fak. Ilmu Komput.*, vol. 9, no. 1, pp. 8–16, 2020.
- [8] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, 2021, doi: 10.30812/bite.v3i1.1110.
- [9] C. Joergensen E Munthe, N. Astuti Hasibuan, and H. Hutabarat, "Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace," *Resolusi Rekayasa Tek. Inform. dan Inf.*, vol. 2, no. 3, pp. 110–115, 2022, doi: 10.30865/resolusi.v2i3.309.
- [10] I. P. D. Suarnatha, I. M. Agus, and O. Gunawan, "Jurnal Computer Science and Information Technology (CoSciTech) manusia," *CoSciTech*, vol. 3, no. 2, pp. 73–80, 2022.
- [11] I. Hadi, L. W. Santoso, and A. N. Tjondrowiguno, "Sistem Rekomendasi Film menggunakan User-based Collaborative Filtering dan K-modes Clustering," *J. Infra*, vol. 3, no. 1, pp. 18–21, 2020, [Online]. Available: <http://publication.petra.ac.id/index.php/teknik-informatika/article/view/9800>
- [12] R. Amelia and D. B. Santoso, "Prediksi Genre Film Dengan Klasifikasi Multi Kelas Sinopsis Menggunakan Jaringan LSTM," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 2, 2023, doi: 10.31539/intecom.v6i2.6961.
- [13] K. A. Nugraha and D. Sebastian, "Pembentukan Dataset Topik Kata Bahasa Indonesia pada Twitter Menggunakan TF-IDF & Cosine Similarity," *J. Tek. Inform. dan Sist. Inf.*, vol. 4, no. 3, pp. 2443–2229, 2018, [Online]. Available: <http://dx.doi.org/10.28932/jutisi.v4i3.862>
- [14] S. Sutejo, "[1] S. Sutejo, 'Pemodelan UML Sistem Informasi Geografis Pasar Tradisional Kota Pekanbaru,' Digit. Zo. J. Teknol. Inf. dan Komun., vol. 7, no. 2, pp. 89–99, 2016. Pemodelan UML Sistem Informasi Geografis Pasar Tradisional Kota Pekanbaru," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 7, no. 2, pp. 89–99, 2016.
- [15] Y. Huliatusuna, "Evaluasi Standar Kemampuan Mahasiswa Terhadap Pembelajaran Metodologi Penelitian Pada Program Studi D-Iv Bidan Pendidik Fakultas Ilmu Kesehatan

Universitas Muhammadiyah Tangerang,” *J. JKFT*, vol. 2, no. 2, p. 71, 2017, doi:
10.31000/jkft.v2i1.695.