

Sistem Rekomendasi Film Berdasarkan Deskripsi Pengguna

Billkaf Syava Al Bana¹

¹Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: *1billkafsyava2004@gmail.com

Abstrak – Pertumbuhan jumlah film dan platform streaming digital telah meningkatkan kompleksitas dalam proses pemilihan film yang sesuai dengan preferensi pengguna. Permasalahan ini semakin menonjol ketika pengguna hanya mampu mendeskripsikan film yang diinginkan secara singkat dan bebas, sehingga pencarian berbasis genre atau kata kunci sering kali menghasilkan rekomendasi yang kurang relevan. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi film berdasarkan deskripsi pengguna dengan menerapkan pendekatan *Content Based Filtering* (CBF). Representasi konten film dan deskripsi pengguna dibangun menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), sedangkan tingkat kesesuaian antar konten diukur menggunakan *Cosine Similarity*. Data yang digunakan dalam penelitian ini bersumber dari dataset terbuka MovieLens yang disediakan oleh GroupLens Research Group dalam format CSV. Evaluasi kinerja sistem menggunakan *intra-list similarity* 0,5148, *diversity* 0,482 dan *coverage* 0,13% menunjukkan bahwa sistem memiliki keseimbangan yang cukup antara kesamaan dan keberagaman rekomendasi meskipun jangkauan rekomendasi masih terbatas. Dengan demikian, sistem yang diusulkan dapat menjadi solusi awal dalam membantu pengguna menemukan film yang sesuai berdasarkan deskripsi yang diberikan.

Kata Kunci — *Content-Based Filtering, Deskripsi, Film, Pengguna, Rekomendasi*

1. PENDAHULUAN

Film merupakan salah satu bentuk media hiburan sekaligus sarana penyampaian informasi yang terus mengalami perkembangan [1] yang signifikan sejalan dengan kemajuan teknologi digital [2] dan semakin luasnya akses masyarakat terhadap platform berbasis daring [3]. Munculnya berbagai layanan *streaming* serta pertumbuhan basis data film yang semakin besar memberikan kebebasan bagi pengguna [4] untuk menentukan pilihan film berdasarkan beragam kriteria seperti *genre*, tema dan deskripsi alur cerita [5]. Akan tetapi, banyaknya alternatif yang tersedia justru memunculkan permasalahan baru, yaitu kesulitan bagi pengguna dalam mengidentifikasi dan memilih film yang benar-benar selaras dengan preferensi serta kebutuhan personal mereka.

Permasalahan dalam menemukan film yang relevan semakin rumit ketika preferensi pengguna bersifat subjektif, dinamis dan tidak selalu dapat dijelaskan hanya melalui klasifikasi *genre* [6]. Dalam banyak kasus, pengguna hanya memiliki gambaran umum atau deskripsi singkat mengenai film yang diharapkan [7], sehingga mekanisme pencarian konvensional berbasis kata kunci kerap menghasilkan rekomendasi yang kurang tepat sasaran [8]. Oleh karena itu, diperlukan suatu pendekatan yang lebih sistematis dan adaptif, yang mampu mengukur tingkat kesesuaian antara konten film dan deskripsi yang diberikan pengguna sebagai landasan dalam menghasilkan rekomendasi yang lebih efektif.

Sebagai upaya untuk menjawab permasalahan tersebut, penelitian ini mengajukan pengembangan sistem rekomendasi film dengan pendekatan *Content-Based Filtering* (CBF) [9] dengan memanfaatkan representasi deskripsi teks deskripsi film menggunakan metode *Term Frequency-Inverse Document* (TF-IDF) [10] serta pengukuran tingkat kemiripan melalui *Cosine Similarity* [11]. Pendekatan ini memungkinkan sistem menghasilkan rekomendasi film yang tersedia objektif dan terukur [12]. Walaupun metode CBF telah banyak diimplementasikan dalam sistem rekomendasi [13], masih terdapat keterbatasan kajian terkait penerapannya secara khusus pada sistem rekomendasi film berbasis deskripsi pengguna yang bersifat ringkas dan bebas, sehingga penelitian ini diharapkan mampu mengisi celah penelitian tersebut.

2. METODE PENELITIAN

Penelitian ini menggunakan metode *Content-Based Filtering* (CBF) dalam membangun sistem rekomendasi film berdasarkan deskripsi pengguna. Metode ini bekerja dengan menganalisis kesesuaian antara karakteristik konten film dan deskripsi yang diberikan pengguna, sehingga rekomendasi yang dihasilkan didasarkan pada tingkat kemiripan konten. Pendekatan CBF dipilih karena mampu memberikan rekomendasi secara personal tanpa bergantung pada data pengguna lain. Adapun tahapan penelitian yang dilakukan sebagai berikut.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini bersumber dari *dataset* terbuka MovieLens yang menyediakan informasi film dalam bentuk terstruktur. *Dataset* tersebut terdiri atas beberapa atribut utama, yaitu *movieId* sebagai identitas unik film, *title* yang merepresentasikan judul film, *genres* yang menggambarkan kategori atau jenis film, *rate* yang menunjukkan penilaian rata-rata pengguna, *tags* yang berisi kata kunci deskriptif, serta *year* yang merepresentasikan tahun rilis film.

Data diperoleh dari sumber terbuka sehingga dapat diakses dan digunakan untuk keperluan pada kelengkapan informasi, konsistensi struktur data, serta relevansinya terhadap pengembangan sistem merekomendasi berbasis Content-Based Filtering yang mengandalkan kesamaan konten sebagai dasar penentuan rekomendasi.

2.2 Pra-pemrosesan Data Teks

Pra-pemrosesan data teks merupakan tahap awal yang dilakukan untuk meningkatkan kualitas data sebelum memasuki proses ekstraksi fitur. Tahapan ini memiliki peran penting karena data teks umumnya masih mengandung variasi penulisan, kata-kata yang kurang relevan, serta struktur yang belum seragam. Pada penelitian ini, pra-pemrosesan diterapkan pada atribut film dengan tujuan menghasilkan representasi teks yang lebih konsisten dan informatif, antara lain:

2.2.1. Case Folding

Case Folding merupakan tahapan pra-pemrosesan teks yang bertujuan mengonversi seluruh karakter huruf menjadi bentuk huruf kecil (*lowercase*). Proses ini dilakukan untuk menghilangkan perbedaan semu yang muncul akibat variasi penggunaan huruf kapital dan non kapital, sehingga kata-kata dengan bentuk leksikal yang sama dapat diperlakukan sebagai satu entitas yang seragam pada tahapan pengolahan berikutnya.

2.2.2. Tokenization

Tokenization merupakan proses pemisahan teks menjadi satuan yang lebih kecil yang dikenal sebagai token, umumnya dalam bentuk kata. Tahapan ini bertujuan untuk memisahkan setiap kata dalam deskripsi film sehingga dapat dianalisis secara individual. Token yang dihasilkan selanjutnya digunakan sebagai dasar dalam proses ekstraksi fitur, mengingat sistem merepresentasikan konten film pada tingkat kata.

2.2.3. Stop Words Removal

Stop words removal merupakan tahapan pra-proses yang dilakukan dengan menghilangkan kata-kata umum yang frekuensinya tinggi tetapi memiliki kontribusi yang rendah terhadap makna teks, seperti kata hubung dan kata depan.

2.2.4. Normalisasi Teks

Normalisasi teks merupakan proses yang bertujuan menyeragamkan bentuk kata sehingga berbagai variasi penulisan dapat direduksi ke dalam satu representasi yang konsisten. Pada tahap ini, teks dibersihkan dari elemen-elemen yang tidak relevan seperti tanda baca, karakter khusus dan simbol yang tidak memiliki kontribusi semantik terhadap isi dokumen.

2.3 TF-IDF

TF-IDF digunakan untuk mengubah data teks hasil pra-pemrosesan menjadi representasi numerik berbentuk vektor. Metode ini mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap seluruh dokumen dalam *dataset*. TF-IDF terdiri dari dua komponen utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF).

Secara sistematis, TF-IDF dirumuskan sebagai berikut:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \dots \dots \dots (1)$$

Dengan:

$$TF(t, d) = \frac{\text{jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{total kata dalam dokumen } d} \dots \dots \dots (2)$$

$$IDF(t) = \log \left(\frac{N}{df(t)} \right) \dots \dots \dots (3)$$

dimana N adalah jumlah total dokumen dan $df(t)$ adalah jumlah dokumen yang mengandung kata t . Melalui pembobotan ini, kata yang sering muncul dalam satu dokumen lain akan memiliki bobot yang lebih tinggi.

2.4 Cosine Similarity

Cosine similarity digunakan untuk mengukur tingkat kemiripan antara dua vektor TF-IDF yaitu vector deskripsi pengguna dan vector atribut film. Metode ini menghitung sudut kosinus antara dua vektor dalam ruang berdimensi tinggi, sehingga fokus pada arah vektor, bukan besarnya.

Rumus cosine similarity dinyatakan sebagai berikut:

$$\text{Cosine Similarity}(A, B) = \frac{A \times B}{\|A\| \times \|B\|} \dots \dots \dots (4)$$

dimana A dan B merupakan vektor TF-IDF, $A \cdot B$ adalah hasil perkalian dot product, serta $\|A\|$ dan $\|B\|$ adalah panjang vector masing-masing. Nilai cosine similarity berada pada rentang 0 hingga 1, dimana nilai yang tinggi menunjukkan kemiripan yang lebih besar.

2.5 Evaluasi Kinerja Sistem

Evaluasi kinerja sistem rekoemendasi dilakukan untuk menilai kualitas rekomendasi yang dihasilkan, terutama ditinjau dari aspek kesamaan, keberagaman, dan cakupan rekomendasi. Pada penelitian ini, proses evaluasi dilakukan dengan menerapkan sejumlah metrik yang sesuai dan relevan dengan karakteristik sistem rekomendasi berbasis konten.

2.5.1. Intra-list Similarity

Intra-list Similarity mengukur tingkat kemiripan antar item dalam satu daftar rekomendasi. Metrik ini digunakan untuk mengetahui sejauh mana film-film yang direkomendasikan memiliki kesamaan konten satu sama lain. Nilai intra-list similarity yang tinggi menunjukkan rekomendasi yang homogen, sedangkan nilai yang lebih rendah mengindikasikan variasi konten yang lebih besar.

2.5.2. Diversity

Diversity merupakan metrik yang digunakan untuk menilai tingkat keberagaman film dalam suatu daftar rekomendasi. Pengukuran ini berperan penting dalam menghindari kecenderungan sistem menghasilkan rekomendasi yang terlalu beragam, sehingga pengguna tetap memperoleh variasi pilihan film yang beragam namun masih selaras dengan deskripsi awal mereka.

2.5.3. Coverage

Coverage digunakan untuk mengukur sejauh mana sistem rekomendasi mampu mencakup keseleuruhan item yang tersedia dalam *dataset*. Pengukuran *coverage* bertujuan untuk menilai kemampuan sistem dalam memanfaatkan koleksi film secara menyeluruh, sehingga rekomendasi yang dihasilkan tidak terbatas pada film-film yang sama secara berulang.

3. HASIL DAN PEMBAHASAN

Pembahasan terhadap hasil penelitian dan pengujian yang diperoleh disajikan dalam bentuk uraian teoritik, baik secara kualitatif maupun kuantitatif. Hasil percobaan sebaiknya ditampilkan dalam berupa grafik atau pun tabel. Untuk grafik dapat mengikuti format untuk diagram dan gambar.

3.1 Hasil Data

Pada tahap pengumpulan data, penelitian ini menggunakan *dataset* film yang diperoleh dari *MovieLens*, yaitu sebuah sumber data terbuka yang disediakan oleh *GroupLens Research Group*. *Dataset* tersebut diperoleh dalam format *Comma Separated Values* (CSV). Berikut sedikit tampilan hasil *dataset*:

Tabel 1. Hasil Data

<i>title</i>	<i>genres</i>	<i>rate</i>	<i>tags</i>	<i>year</i>
<i>Toy Story</i>	<i>Adventure</i> <i>Animation Children</i> <i>Comedy Fantasy</i>	3.89	<i>animation</i> <i>friendship toys</i> <i>animation Disney</i> <i>Pixar toys CGI</i> <i>classic disney pixar</i> <i>lots of heart Tom</i> <i>Hanks</i>	1995
<i>The Amazing Spiderman</i>	<i>Action Adventure</i> <i>SciFi IMAX</i>	3.24	<i>action Andrew</i> <i>Garfield Bad</i> <i>Adaptation based</i> <i>on a comic comic</i> <i>book Denis Leary</i> <i>Emma Stone</i> <i>Genetic</i> <i>engineering Martin</i> <i>Sheen Marvel</i> <i>nerds</i>	2012
<i>2 Fast 2 Furious</i>	<i>Action Crime</i> <i>Thriller</i>	2.64	<i>cars crappy sequel</i> <i>tattoo cars dumb</i> <i>Eva Mendes miami</i> <i>Paul Walker street</i> <i>race tuning Cars</i> <i>street race street</i> <i>race not as good</i>	2003
<i>Finding Nemo</i>	<i>Adventure</i> <i>Animation Children</i> <i>Comedy</i>	3.82	<i>fatherson</i> <i>relationship</i> <i>overrated</i> <i>predictable</i> <i>protective father</i> <i>unoriginal</i> <i>animated</i> <i>animation Disney</i> <i>family fatherson</i> <i>relationship fish</i> <i>friendship</i> <i>heartwarming</i> <i>ocean Pixar</i> <i>shortterm memory</i> <i>loss talking</i> <i>animals</i>	2003

3.2 Pra-pemrosesan Data Teks

3.2.1. Case Folding

Pada *case folding*, seluruh teks pada atribut film diubah menjadi huruf kecil (*lowercase*) Proses ini diterapkan pada data hasil pengumpulan untuk menghilangkan perbedaan penulisan akibat penggunaan huruf kapital yang tidak konsisten.

Tabel 2. Hasil *Case Folding*

<i>title</i>	<i>genres</i>	<i>rate</i>	<i>tags</i>	<i>year</i>
<i>the amazing spiderman</i>	<i>action adventure scifi imax</i>	3.24	<i>action andrew garfield bad adaptation based on a comic comic book denis leary emma stone genetic engineering martin sheen marvel nerds</i>	2012

3.2.2. Tokenization

Pada tahap *tokenization*, teks hasil *case folding* dipecah menjadi unit-unit kata (token) berdasarkan pemisah teretntu, seperti spasi. Proses ini bertujuan untuk memisahkan setiap kata dalam atribut film, sehingga dapat dianalisis secara individual pada tahap pengolahan selanjutnya.

Tabel 3. Hasil *Tokenization*

<i>title</i>	<i>genres</i>	<i>rate</i>	<i>tags</i>	<i>year</i>
<i>[the, amazing, spiderman]</i>	<i>[action, adventure, scifi, imax]</i>	[3.24]	<i>[action, andrew, garfield, bad, adaptation, based, on, a, comic, comic, book, denis, leary, emma, stone, genetic, engineering, martin, sheen, marvel, nerds]</i>	[2012]

3.2.3. Stop Words Removal

Pada tahap *stop words removal*, kata-kata umum yang sering muncul namun tidak memiliki makna signifikan terhadap atribut film dihapus dari hasil tokenisasi. *Stop words removal* dilakukan untuk mengurangi noise pada data teks sehingga hanya kata-kata yang relevan yang dipertahankan sebagai representasi konten film. Pada tabel 4 menunjukkan beberapa kata yang tidak penting dihapus seperti *the*, *on*, *a* dan *based*.

Tabel 4. Hasil *Stop Words Removal*

<i>title</i>	<i>genres</i>	<i>rate</i>	<i>tags</i>	<i>year</i>
<i>[amazing, spiderman]</i>	<i>[action, adventure, scifi, imax]</i>	[3.24]	<i>[action, andrew, garfield, bad, adaptation, comic, comic, book, denis, leary, emma, stone, genetic, engineering, martin, sheen, marvel, nerds]</i>	[2012]

3.2.4. Normalisasi Teks

Selanjutnya normalisasi teks, dilakukan penyamaan bentuk kata untuk mengurangi variasi penulisan yang memiliki makna serupa serta menghilangkan tanda baca yang tidak penting. Normalisasi teks bertujuan agar kata-kata dalam atribut film menjadi lebih konsisten sehingga representasi teks yang dihasilkan lebih

akurat untuk tahap analisis selanjutnya. Pada tabel 5 menunjukkan beberapa kata memiliki makna yang sama seperti *action* dan *comic*.

Tabel 5. Hasil Normalisasi Teks

<i>title</i>	<i>genres</i>	<i>rate</i>	<i>tags</i>	<i>year</i>
[<i>amazing, spiderman</i>]	[<i>action, adventure, scifi, imax</i>]	[3.24]	[<i>andrew, garfield, bad, adaptation, comic, book, denis, leary, emma, stone, genetic, engineering, martin, sheen, marvel, nerds</i>]	[2012]

3.3 TF-IDF

Pada tahap TF-IDF, data teks hasil pra=pemrosesan diubah menjadi representasi numerik dalam bentuk vektor. Proses ini dilakukan dengan memberikan bobot pada setiap kata berdasarkan tingkat kemunculannya dalam satu deskripsi film dan tingkat kelangkaannya pada keseluruhan *dataset*, sehingga yang lebih representatif memiliki pengaruh besar dalam sistem rekomendasi. Berikut adalah perhitungan sederhana dari TF-IDF

Misal seorang pengguna memberikan deskripsi berupa “berikan saya rekomendasi film bergenre action adventure scifi yang diperankan oleh andrew garfield dan emma stone bertemakan superhero marvel”

Tabel 6. Perhitungan TF

Deskripsi (D0)		<i>Toy Story</i> (D1)		<i>The Amazing Spiderman</i> (D2)		<i>2 Fast 2 Furious</i> (D3)		<i>Finding Nemo</i> (D4)	
<i>action</i>	1/8 = 0,125			<i>action</i>	1/16 = 0,062	<i>action</i>	1/14 = 0,071		
<i>adventure</i>	1/8 = 0,125	<i>adventure</i>	1/11 = 0,090	<i>adventure</i>	1/16 = 0,062			<i>adventure</i>	1/11 = 0,090
<i>scifi</i>	1/8 = 0,125			<i>scifi</i>	1/16 = 0,062				
<i>andrew</i>	1/8 = 0,125			<i>andrew</i>	1/16 = 0,062				
<i>garfield</i>	1/8 = 0,125			<i>garfield</i>	1/16 = 0,062				
<i>emma</i>	1/8 = 0,125			<i>emma</i>	1/16 = 0,062				
<i>stone</i>	1/8 = 0,125			<i>stone</i>	1/16 = 0,062				
<i>marvel</i>	1/8 = 0,125			<i>marvel</i>	1/16 = 0,062				

Tabel 7. Perhitungan IDF

Kata	Muncul di dokumen	DF	IDF
<i>action</i>	D0, D2, D3	3	0,426
<i>adventure</i>	D0, D1, D2, D4	4	0,301
<i>scifi</i>	D0, D2	2	0,602
<i>andrew</i>	D0, D2	2	0,602
<i>garfield</i>	D0, D2	2	0,602
<i>emma</i>	D0, D2	2	0,602
<i>stone</i>	D0, D2	2	0,602
<i>marvel</i>	D0, D2	2	0,602

Tabel 8. Perhitungan TF × IDF

Kata	Deskripsi (D0)	<i>Toy Story</i> (D1)	<i>The Amazing Spiderman</i> (D2)	<i>2 Fast 2 Furious</i> (D3)	<i>Finding Nemo</i> (D4)
action	0,053		0,026	0,030	
adventure	0,037	0,27	0,018		0,27
scifi	0,752		0,037		
andrew	0,752		0,037		
garfield	0,752		0,037		
emma	0,752		0,037		
stone	0,752		0,037		
marvel	0,752		0,037		

3.4 Cosine Similarity

Pada tahap *cosine similarity*, dilakukan perhitungan tingkat kemiripan antara vektor TF-IDF deskripsi pengguna dan vektor TF-IDF atribut film. Nilai kemiripan ini digunakan sebagai dasar untuk menentukan tingkat relevansi film, dimana film dengan nilai *cosine similarity* tertinggi dianggap paling sesuai dan direkomendasikan kepada pengguna.

Tabel 9. Perhitungan *Cosine Similarity*

Pasangan	Skor
D0 – D1	0,034
D0 – D2	0,999
D0 – D3	0,003
D0 – D4	0,034

Nilai *cosine similarity* tertinggi diperoleh oleh pasangan D0-D2 dengan nilai 0,999 yang menunjukkan bahwa film *The Amazing Spiderman* memiliki tingkat kemiripan paling tinggi dengan deskripsi pengguna. Dengan demikian, film tersebut dinilai paling relevan dan layak direkomendasikan.

3.5 Evaluasi Kinerja Sistem

3.5.1. Intra-list Similarity

Pada tahap *intra-list similarity*, dilakukan pengukuran tingkat kemiripan antar film dalam satu daftar rekomendasi. Pengukuran ini bertujuan untuk mengetahui sejauh mana film-film yang direkomendasikan memiliki kesamaan konten berdasarkan representasi vektor yang digunakan. Dalam deskripsi “berikan saya rekomendasi film bergenre *action adventure* bertemakan *marvel superhero*” menampilkan rekomendasi seperti ini.

Tabel 9. Hasil Rekomendasi Film dengan Deskripsi Berbeda

title	genres	year	rate	similarity
<i>LEGO Marvel SpiderMan Vexed By Venom</i>	Action Animation	2019	2.90	0.548238
<i>Avengers Age of Ultron</i>	Action Adventure SciFi	2015	3.50	0.490600
<i>Marvel OneShot Item 47</i>	Action Fantasy SciFi	2012	3.27	0.454975
<i>Captain America The First Avenger</i>	Action Adventure SciFi Thriller War	2011	3.48	0.453769
<i>Thor</i>	Action Adventure Drama Fantasy IMAX	2011	3.32	0.448696

Dari nilai *similarity* antar film yang diperoleh, dapat disimpulkan bahwa sistem rekomendasi menghasilkan tingkat kemiripan konten yang cukup konsisten dan berada pada kategori menengah. Rentang nilai tersebut menunjukkan bahwa film-film yang direkomendasikan memiliki kesamaan deskripsi yang relatif signifikan. Dan nilai *intra-list similarity* yang diperoleh yaitu 0,5148 yang memperkuat temuan

tersebut, yang mengindikasikan bahwa rata-rata kemiripan antar item dalam satu daftar rekomendasi berada sedikit diatas nilai tengah skala kemiripan.

3.5.2. Diversity

Lalu pada nilai *diversity* nya diperoleh sebesar 0,4852 menunjukkan bahwa daftar rekomendasi yang dihasilkan memiliki tingkat keberagaman yang sedang. Nilai ini mengindikasikan bahwa meskipun film-film yang direkomendasikan masih memiliki keterkaitan konten yang cukup kuat, sistem tidak sepenuhnya menyajikan item yang seragam atau sangat mirip satu sama lain dan inipun wajar karena berbeda deskripsi berbeda juga nilai nya.

3.5.3. Coverage

Nilai *coverage* yang diperoleh yaitu sebesar 0,13% yang berarti 63 dari 50,154 film dalam *dataset* berhasil terakomodasi dalam hasil rekomendasi, menunjukkan bahwa jangkauan sistem rekomendasi masih terbatas. Temuan ini mengindikasikan bahwa sistem cenderung merekomendasikan dari subset kecil dalam keseluruhan koleksi yang tersedia.

4. SIMPULAN

Setelah melalui seluruh tahapan penelitian yang meliputi pengumpulan data, pra-pemrosesan data teks, ekstraksi fitur menggunakan TF-IDF, perhitungan tingkat kemiripan dengan *cosine similarity* serta evaluasi kinerja sistem yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut:

1. Sistem rekomendasi film berbasis deskripsi pengguna berhasil dibangun menggunakan pendekatan CBF dengan metode TF-IDF dan *cosine similarity* sebagai pengukuran kemiripan konten film.
2. Tahapan pra-pemrosesan data teks, meliputi, *case folding*, *tokenization*, *stop words removal* dan normalisasi teks, terbukti mampu meningkatkan konsistensi dan kualitas representasi data sebelum dilakukan ekstraksi fitur.
3. Hasil perhitungan *cosine similarity* menunjukkan bahwa sistem mampu mengidentifikasi film dengan tingkat kesesuaian tinggi terhadap deskripsi pengguna sehingga rekomendasi yang dihasilkan bersifat relevan dan terukur.
4. Evaluasi kinerja sistem menunjukkan nilai *intra-list similarity* dan *diversity* berada pada kategori sedang, yang mengindikasikan bahwa sistem mampu menjaga keseimbangan antara kesamaan dan keberagaman rekomendasi.
5. Nilai *coverage* yang masih rendah menunjukkan bahwa sistem cenderung merekomendasikan film dari subset tertentu dalam *dataset*, sehingga pemanfaatan koleksi film secara keseluruhan belum optimal.

5. SARAN

Berdasarkan keterbatasan yang ditemukan dalam penelitian ini, pengembangan penelitian selanjutnya disarankan untuk mengkombinasikan metode *Content-Based Filtering* dengan pendekatan lain, seperti *Collaborative Filtering*, guna meningkatkan jangkauan serta variasi rekomendasi yang dihasilkan. Selain itu, pengayaan fitur teks berupa penambahan sinopsis film perlu dipertimbangkan agar guna memperkaya representasi konten dan meningkatkan tingkat kesesuaian antara deskripsi pengguna dan film yang direkomendasikan secara lebih optimal

DAFTAR PUSTAKA

- [1] M. B. Mustofa *et al.*, “FUNGSI KOMUNIKASI MASSA DALAM FILM,” *AT-TAWASUL*, vol. 2, no. 1, pp. 1–8, Aug. 2022, doi: 10.51192/JA.V2I1.324.
- [2] E. A. Sahrul and K. Nuringsih, “PERAN E-COMMERCE, MEDIA SOSIAL DAN DIGITAL TRANSFORMATION UNTUK PENINGKATAN KINERJA BISNIS UMKM,” *Jurnal Muara Ilmu Ekonomi dan Bisnis*, vol. 7, no. 2, pp. 286–299, Oct. 2023, doi: 10.24912/jmieb.v7i2.23293.
- [3] J. Dwi Sahputra Amory, M. Mudo, and U. Muhammadiyah Mamuju, “Transformasi Ekonomi Digital dan Evolusi Pola Konsumsi: Tinjauan Literatur tentang Perubahan Perilaku Belanja di Era Internet,” *Jurnal Minfo Polgan*, vol. 14, no. 1, pp. 28–37, Feb. 2025, doi: 10.33395/JMP.V14I1.14608.
- [4] M. Y. Patawari and D. R. Sugiyanto, “Pergeseran Paradigma Distribusi Film Pada Masa Pandemi Covid-19 dan Prediksi Eksistensi Bioskop Pascapandemi Covid-19,” *Urban: Jurnal Seni Urban*, vol. 5, no. 2, pp. 79–98, Oct. 2021, doi: 10.52969/JSU.V5I2.56.
- [5] K. Genre *et al.*, “Klasifikasi Genre Film Terpopuler Bulanan Menggunakan Algoritma Naive Bayes Berbasis Data Penayangan,” *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 9, no. 3, pp. 2183–2190, Jul. 2025, doi: 10.29407/40KX4V59.
- [6] R. Insan Pratama Siagian, N. Khoiriah, S. Audy Priscilia, M. Raffi Akbar Tanjung, and A. Perdana, “PENERAPAN MACHINE LEARNING UNTUK REKOMENDASI FILM BERDASARKAN PREFERENSI PENGGUNA,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 5658–5662, May 2025, doi: 10.36040/JATI.V9I4.13884.
- [7] A. Suhendra and A. Ahmad, “Perancangan Film Pendek Drama tentang Digital Addiction Berjudul Digislave,” *Jurnal Desain*, vol. 10, no. 2, p. 331, Jan. 2023, doi: 10.30998/JD.V10I2.14332.
- [8] D. Velamentosa, E. Zuliarso, and J. Raya Tri Lomba Juang, “SISTEM REKOMENDASI FILM MENGGUNAKAN METODE CONTENT-BASED FILTERING,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, pp. 2918–2922, Mar. 2025, doi: 10.36040/JATI.V9I2.13251.
- [9] S. Noorlima Yanti, E. Rihyanti, J. Margonda Raya, and D. Jawa Barat, “Penerapan Rest API untuk Sistem Informasi Film secara Daring,” *Jurnal Informatika Universitas Pamulang*, vol. 6, no. 1, pp. 195–201, Mar. 2021, doi: 10.32493/INFORMATIKA.V6I1.10033.
- [10] A. H. Azmi, I. Naufal, R. Mahardika, A. Shaktika, A. Prasetya, and S. Puspita, “Sistem Rekomendasi Film Berbasis Konten Menggunakan Teknik Cosine Similarity dan TF-IDF,” *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, vol. 1, no. 4, pp. 2135–2145, Sep. 2024, doi: 10.32672/MISTER.V1I4.2206.
- [11] F. Yuniardini and T. Widiyaningtyas, “Analisis Perbandingan Pearson Correlation dan Cosine Similarity pada Rekomendasi Musik berbasis Collaborative Filtering,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 555–564, Dec. 2024, doi: 10.29408/edumatic.v8i2.27781.
- [12] A. Dian Anggi Kusuma, A. Pramudwiatmoko, M. Teknologi Informasi, and A. Info, “Sistem Rekomendasi Pemilihan Film Menggunakan Metode Collaborative Filtering dan Content-Based Filtering,” *INTEK : Jurnal Informatika dan Teknologi Informasi*, vol. 8, no. 1, pp. 88–97, May 2025, doi: 10.37729/INTEK.V8I1.6286.
- [13] A. Pranata and F. Sulianta, “Implementasi dan Analisis Algoritma Content-Based Filtering Pada Sistem Rekomendasi Produk Tas pada Basis Data MySQL,” *Jurnal Ilmiah Global Education*, vol. 6, no. 3, pp. 1419–1444, Aug. 2025, doi: 10.55681/JIGE.V6I3.4017.