

Perancangan Sistem Rekomendasi Film Indonesia Menggunakan Content-Based Filtering

Ikhsan Nur Purwanto^{1*}, Muhammad Zamharir Rioga P², Galuh Wisma Safitri³, Nur Elsa Widyaningrum⁴, Joevan Maulana Florian⁵

¹Teknik Informatika/Fakultas Ilmu Komputer

Afiliasi Universitas Duta Bangsa

^{1*}230103024@mhs.udb.ac.id (penulis korespondensi)

²Teknik Informatika/Fakultas Ilmu Komputer

Afiliasi Universitas Duta Bangsa

²230103031@mhs.udb.ac.id

³Teknik Informatika/Fakultas Ilmu Komputer

Afiliasi Universitas Duta Bangsa

³230103022@mhs.udb.ac.id

⁴Teknik Informatika/Fakultas Ilmu Komputer

Afiliasi Universitas Duta Bangsa

⁴230103033@mhs.udb.ac.id

⁵Teknik Informatika/Fakultas Ilmu Komputer

Afiliasi Universitas Duta Bangsa

⁵230103139@mhs.udb.ac.id

Abstrak— Perkembangan layanan hiburan digital menyebabkan jumlah pilihan film semakin banyak, sehingga pengguna sering mengalami kesulitan dalam menentukan film yang sesuai dengan minatnya. Penelitian ini bertujuan untuk merancang sistem rekomendasi film berdasarkan kemiripan konten dan metadata film menggunakan metode Content-Based Filtering. Data yang digunakan terdiri dari 20 data film dengan atribut judul film, tahun, genre, sinopsis singkat, kata kunci, sutradara, dan pemain utama. Lima variabel utama yang digunakan dalam proses rekomendasi adalah genre, sinopsis, kata kunci, sutradara, dan pemain utama. Tahapan perancangan meliputi pengumpulan data, preprocessing, pembobotan kata menggunakan Term Frequency-Inverse Document Frequency (TF-IDF), normalisasi vektor, perhitungan kemiripan antarfilm menggunakan Cosine Similarity, dan perancangan antarmuka. Hasil penelitian menunjukkan bahwa pada film acuan “Dilan 1990”, film yang direkomendasikan adalah “Dilan 1991” dengan nilai similarity sebesar 0,568. Sementara itu, pada film acuan “KKN di Desa Penari”, film yang direkomendasikan adalah “Badarawuhi di Desa Penari” dengan nilai similarity sebesar 0,304. Hasil tersebut menunjukkan bahwa rancangan metode dapat menggambarkan proses rekomendasi film berdasarkan kesamaan konten dan metadata. Namun, penelitian ini masih terbatas pada tahap perancangan dan menggunakan dataset kecil, sehingga hasilnya bersifat ilustratif dan belum dapat membuktikan efektivitas sistem secara empiris.

Kata kunci— Sistem Rekomendasi, Film, Content-Based Filtering, TF-IDF, Cosine Similarity

Abstract— The rapid development of digital entertainment services has increased the number of available movie choices, making it difficult for users to select movies that match their interests. This study aims to design a movie recommendation system based on the similarity of movie content and metadata using the Content-Based Filtering method. The dataset used in this study consists of 20 movie items with attributes including movie title, year, genre, short synopsis, keywords, director, and main cast. The five main variables used in the recommendation process are genre, synopsis, keywords, director, and main cast. The design stages include data collection, preprocessing, word weighting using Term Frequency-Inverse Document Frequency (TF-IDF), vector normalization, similarity calculation between movies using Cosine Similarity, and user interface design. The results show that for the reference movie “Dilan 1990”, the recommended movie is “Dilan 1991” with a similarity score of 0.568. Meanwhile, for the reference movie “KKN di Desa Penari”, the recommended movie is “Badarawuhi di Desa Penari” with a similarity score of 0.304. These results indicate that the proposed method design can illustrate the movie recommendation process based on similarities in content and metadata. However, this study is limited to the design stage and uses a small dataset, so the results are illustrative and cannot yet prove the empirical effectiveness of the system.

Keywords— Recommendation System, Movie, Content-Based Filtering, TF-IDF, Cosine Similarity.

I. PENDAHULUAN

Perkembangan teknologi informasi telah memberikan pengaruh besar terhadap cara masyarakat mengakses hiburan digital, salah satunya melalui layanan film. Banyaknya pilihan film dengan berbagai genre, tema cerita, tokoh,

dan karakteristik konten membuat pengguna sering mengalami kesulitan dalam menentukan film yang sesuai dengan minatnya. Kondisi tersebut menunjukkan perlunya sistem yang dapat membantu pengguna memperoleh rekomendasi film secara lebih terarah

berdasarkan kesesuaian konten film dengan preferensi pengguna [1].

Sistem rekomendasi merupakan salah satu solusi yang dapat digunakan untuk membantu pengguna dalam memilih item yang sesuai dengan kebutuhannya [2]. Pada bidang film, sistem rekomendasi dapat digunakan untuk memberikan daftar film yang memiliki kemiripan dengan film yang pernah disukai atau dipilih oleh pengguna [3]. Salah satu metode yang banyak digunakan adalah Content-Based Filtering, yaitu metode rekomendasi yang bekerja dengan membandingkan karakteristik atau atribut dari suatu item. Pada rekomendasi film, atribut yang dapat digunakan antara lain genre, sinopsis, kata kunci, sutradara, pemain utama, maupun informasi lain yang menggambarkan isi dan karakteristik film [4].

Beberapa penelitian sebelumnya menunjukkan bahwa metode Content-Based Filtering dapat diterapkan dalam sistem rekomendasi film. Penelitian Fajriansyah dkk [4]. menerapkan Content-Based Filtering untuk memberikan rekomendasi film berdasarkan kemiripan konten film. Penelitian lain menggunakan sinopsis film sebagai dasar rekomendasi dengan menerapkan TF-IDF dan Cosine Similarity untuk menghitung kemiripan antarfilm. Selain itu, penelitian Fuady dkk [5]. juga memanfaatkan genre dan sinopsis sebagai atribut rekomendasi film berbasis Content-Based Filtering. Penelitian Santoso dkk [6]. mengembangkan prototipe sistem rekomendasi film Indonesia dengan pendekatan Content-Based Filtering dan Vector Space Model. Berdasarkan beberapa penelitian tersebut, metode Content-Based Filtering dinilai sesuai digunakan pada perancangan sistem rekomendasi film karena dapat memanfaatkan atribut konten dan metadata film sebagai dasar pencarian kemiripan [7].

Penelitian ini berfokus pada perancangan sistem rekomendasi film berdasarkan lima variabel, yaitu genre, sinopsis, kata kunci, sutradara, dan pemain utama menggunakan metode Content-Based Filtering. Genre digunakan untuk mengetahui kategori film, sinopsis digunakan untuk menggambarkan isi cerita, kata kunci digunakan untuk mewakili

tema utama, sutradara digunakan sebagai informasi pembuat film, sedangkan pemain utama digunakan untuk menunjukkan tokoh atau aktor yang berperan penting dalam film. Teknik TF-IDF digunakan untuk memberikan bobot pada kata atau istilah yang terdapat dalam atribut film, sedangkan Cosine Similarity digunakan untuk menghitung tingkat kemiripan antarfilm [8].

Data yang digunakan dalam penelitian ini terdiri dari 20 data film yang disusun sebagai data item dengan atribut judul film, tahun, genre, sinopsis singkat, kata kunci, sutradara, dan pemain utama. Dari atribut tersebut, variabel yang digunakan dalam proses rekomendasi adalah genre, sinopsis, kata kunci, sutradara, dan pemain utama. Penelitian ini hanya dibatasi pada tahap perancangan, meliputi perancangan proses, perancangan metode, perancangan data, dan perancangan antarmuka sistem.

Tujuan dari penelitian ini adalah menghasilkan rancangan sistem rekomendasi film yang dapat memberikan rekomendasi berdasarkan kemiripan konten dan metadata film. Hasil dari penelitian ini diharapkan dapat menjadi dasar pengembangan aplikasi rekomendasi film berbasis web yang membantu pengguna menemukan film sesuai dengan minat dan karakteristik film yang dipilih.

II. METODOLOGI PENELITIAN

Metodologi penelitian ini menjelaskan tahapan yang dilakukan dalam merancang sistem rekomendasi film berdasarkan genre, sinopsis, kata kunci, sutradara, dan pemain utama menggunakan metode Content-Based Filtering. Penelitian ini berfokus pada tahap perancangan sistem, sehingga hasil yang dihasilkan meliputi rancangan proses, rancangan metode, rancangan data, dan rancangan antarmuka sistem rekomendasi.

A. Tahapan Penelitian

Tahapan penelitian merupakan langkah-langkah yang dilakukan dalam proses perancangan sistem rekomendasi film. Tahapan ini disusun agar proses penelitian berjalan secara terarah, mulai dari pengumpulan data hingga

perancangan antarmuka sistem. Adapun tahapan penelitian yang dilakukan meliputi:

1. Pengumpulan Data
2. Preprocessing
3. Content-Based Filtering
4. Perancangan Antarmuka

Tahap pertama adalah pengumpulan data. Pada tahap ini dilakukan pengumpulan data film yang akan digunakan sebagai data item dalam sistem rekomendasi. Data yang digunakan terdiri dari 20 data film dengan atribut judul film, tahun, genre, sinopsis singkat, kata kunci, sutradara, dan pemain utama. Atribut tersebut dipilih karena dapat menggambarkan karakteristik film, baik dari sisi isi cerita maupun metadata film.

Tahap kedua adalah preprocessing. Pada tahap ini, lima variabel utama, yaitu genre, sinopsis, kata kunci, sutradara, dan pemain utama digabungkan menjadi satu dokumen teks untuk setiap film. Selanjutnya, data teks diproses melalui beberapa tahapan, seperti case folding, tokenizing, stopword removal, dan stemming [9]. Tahapan preprocessing dilakukan agar data teks menjadi lebih terstruktur dan siap digunakan dalam proses pembobotan kata.

Tahap ketiga adalah penerapan metode Content-Based Filtering. Pada tahap ini, sistem membandingkan karakteristik antarfilm berdasarkan fitur teks yang telah diproses sebelumnya. Pembobotan kata dilakukan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) untuk menghasilkan bobot pada setiap kata atau fitur [10]. Setelah itu, tingkat kemiripan antarfilm dihitung menggunakan Cosine Similarity [11]. Film dengan nilai kemiripan tertinggi akan ditampilkan sebagai hasil rekomendasi [12].

Tahap keempat adalah perancangan antarmuka. Pada tahap ini dilakukan perancangan tampilan sistem rekomendasi film yang terdiri dari halaman daftar film, halaman detail film, dan halaman hasil rekomendasi. Perancangan antarmuka bertujuan untuk memberikan gambaran alur interaksi pengguna dengan sistem, mulai dari memilih film hingga memperoleh daftar rekomendasi film yang memiliki kemiripan tertinggi.

B. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data item film yang disusun untuk kebutuhan perancangan sistem rekomendasi. Jumlah data yang digunakan sebanyak 20 film. Setiap data film memiliki beberapa atribut, yaitu judul film, tahun rilis, genre, sinopsis singkat, kata kunci, sutradara, dan pemain utama. Atribut genre, sinopsis, kata kunci, sutradara, dan pemain utama digunakan sebagai dasar dalam proses rekomendasi karena kelima atribut tersebut dapat menggambarkan karakteristik isi, tema, gaya, dan tokoh yang terdapat pada film.

Contoh struktur data film yang digunakan dalam penelitian ini dapat dilihat pada Tabel I.

TABEL I. STRUKTUR DATA FILM

No	Atribut	Keterangan
1	Judul Film	Nama film yang digunakan sebagai data item
2	Tahun	Tahun rilis film
3	Genre	Kategori film, seperti drama, komedi, romantis, horor, atau keluarga
4	Sinopsis Singkat	Ringkasan cerita film yang menggambarkan isi utama film
5	Kata Kunci	Kata penting yang menggambarkan tema atau karakteristik film
6	Sutradara	Nama pembuat atau pengarah film
7	Pemain Utama	Nama aktor atau tokoh utama yang berperan dalam film

C. Preprocessing Data

Sebelum data diproses menggunakan metode TF-IDF, atribut genre, sinopsis, kata kunci, sutradara, dan pemain utama digabungkan menjadi satu kolom fitur teks. Penggabungan ini bertujuan agar sistem dapat membaca keseluruhan informasi yang menggambarkan karakteristik film. Genre digunakan untuk menunjukkan kategori film, sinopsis digunakan untuk menjelaskan isi cerita, kata kunci digunakan untuk mewakili tema utama, sutradara digunakan untuk menunjukkan pengarah film, sedangkan pemain utama digunakan untuk menggambarkan tokoh atau aktor yang berperan penting dalam film [5].

Setelah atribut digabungkan, dilakukan preprocessing teks agar data lebih siap digunakan dalam proses pembobotan [9]. Tahapan preprocessing yang dilakukan meliputi case folding, yaitu mengubah seluruh huruf menjadi huruf kecil; tokenizing, yaitu memecah kalimat

menjadi kata-kata; stopword removal, yaitu menghapus kata umum yang tidak memiliki makna penting; dan stemming, yaitu mengubah kata berimbuhan menjadi bentuk kata dasar [9]. Hasil preprocessing kemudian digunakan sebagai masukan pada proses pembobotan TF-IDF [8].

D. Content-Based Filtering

Metode yang digunakan dalam penelitian ini adalah Content-Based Filtering [4]. Metode ini bekerja dengan membandingkan karakteristik antaritem berdasarkan atribut yang dimiliki [2]. Pada penelitian ini, item yang direkomendasikan adalah film. Sistem akan memberikan rekomendasi film yang memiliki kemiripan konten dan metadata dengan film yang dipilih oleh pengguna [7]. Kemiripan tersebut dihitung berdasarkan lima variabel utama, yaitu genre, sinopsis, kata kunci, sutradara, dan pemain utama [3].

Content-Based Filtering dipilih karena metode ini tidak membutuhkan data rating dari banyak pengguna. Sistem cukup menggunakan data karakteristik film untuk menghasilkan rekomendasi [1]. Hal ini sesuai dengan kebutuhan penelitian yang menggunakan data item sebanyak 20 film sebagai dasar perancangan sistem.

Pembobotan TF-IDF

TF-IDF atau Term Frequency-Inverse Document Frequency digunakan untuk memberikan bobot pada setiap kata yang terdapat dalam data film [10]. Term Frequency menunjukkan seberapa sering suatu kata muncul dalam satu dokumen, sedangkan Inverse Document Frequency menunjukkan seberapa penting kata tersebut dibandingkan dengan seluruh dokumen [8]. Dalam penelitian ini, setiap film dianggap sebagai satu dokumen, sedangkan gabungan genre, sinopsis, kata kunci, sutradara, dan pemain utama menjadi isi dokumen.

Secara umum, pembobotan TF-IDF dapat dituliskan sebagai berikut:

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

Keterangan:

TF-IDF(t,d) adalah bobot kata t pada dokumen d, TF(t,d) adalah frekuensi kemunculan kata t pada dokumen d, dan IDF(t) adalah nilai kepentingan kata t terhadap seluruh dokumen.

Setelah proses pembobotan menggunakan TF-IDF, setiap film direpresentasikan dalam bentuk vektor numerik. Vektor TF-IDF tersebut kemudian dinormalisasi menggunakan normalisasi L2. Normalisasi L2 dilakukan untuk menyamakan skala panjang vektor setiap dokumen film, karena setiap film dapat memiliki jumlah kata dan panjang teks yang berbeda. Dengan normalisasi ini, panjang vektor setiap dokumen menjadi bernilai 1, sehingga proses perhitungan Cosine Similarity dapat membandingkan arah kemiripan antarvektor, bukan dipengaruhi oleh panjang dokumen.

Setelah vektor TF-IDF dinormalisasi, tahap berikutnya adalah menghitung kemiripan antarfilm menggunakan Cosine Similarity. Nilai Cosine Similarity diperoleh dari perbandingan antara dot product vektor film acuan dan film pembanding dengan hasil perkalian panjang kedua vektor. Pada vektor yang telah melalui normalisasi L2, nilai panjang vektor menjadi 1, sehingga hasil perhitungan similarity setara dengan dot product antarvektor yang telah dinormalisasi.

E. Perhitungan Cosine Similarity

Setelah setiap film memiliki bobot TF-IDF, tahap berikutnya adalah menghitung tingkat kemiripan antarfilm menggunakan Cosine Similarity [11]. Cosine Similarity digunakan untuk mengukur kedekatan antara dua dokumen berdasarkan vektor bobot kata [12]. Semakin tinggi nilai Cosine Similarity, maka semakin mirip karakteristik kedua film tersebut [11].

Rumus Cosine Similarity dapat dituliskan sebagai berikut:

Cosine Similarity

$$(A, B) = (A \cdot B) / (\|A\| \|B\|)$$

Keterangan:

A dan B adalah vektor bobot TF-IDF dari dua film yang dibandingkan, $A \cdot B$ adalah hasil perkalian dot product antara vektor A dan B, sedangkan $\|A\|$ dan $\|B\|$ adalah panjang vektor masing-masing film.

F. Perancangan Proses Rekomendasi

Perancangan proses rekomendasi pada penelitian ini dimulai dari penginputan data film sebanyak 20 data film ke dalam sistem. Setiap data film memiliki beberapa atribut, yaitu genre,

sinopsis, kata kunci, sutradara, dan pemain utama. Kelima variabel tersebut kemudian diambil dan digabungkan menjadi satu dokumen teks yang mewakili karakteristik masing-masing film. Setelah itu, sistem melakukan preprocessing teks yang meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming* agar data teks lebih terstruktur dan siap digunakan pada tahap perhitungan.

Tahap selanjutnya adalah pembobotan kata menggunakan metode TF-IDF untuk memperoleh representasi numerik dari setiap dokumen film [10]. Setelah seluruh data film diproses, pengguna memilih salah satu film yang tersedia pada sistem sebagai input awal rekomendasi. Sistem kemudian menghitung tingkat kemiripan antara film yang dipilih dengan film lainnya menggunakan *Cosine Similarity* [11]. Nilai kemiripan yang diperoleh akan diurutkan dari nilai tertinggi ke nilai terendah, sehingga sistem dapat menampilkan daftar film yang paling relevan sebagai hasil rekomendasi [13].

Secara keseluruhan, alur proses rekomendasi dalam penelitian ini meliputi input data film, pengambilan atribut film, penggabungan lima variabel menjadi dokumen teks, preprocessing teks, pembobotan TF-IDF, pemilihan film oleh pengguna, perhitungan kemiripan menggunakan *Cosine Similarity*, pengurutan nilai *similarity*, dan penampilan daftar rekomendasi film. Alur tersebut ditunjukkan pada Gambar 1.

Gambar 1. Flowchart Alur Proses Sistem Rekomendasi Film



Gambar 1. Flowchart Alur Proses Sistem Rekomendasi Film

H. Perancangan Antarmuka Sistem

Perancangan antarmuka sistem dilakukan untuk memberikan gambaran mengenai tampilan dan alur interaksi pengguna dalam sistem rekomendasi film. Antarmuka yang dirancang terdiri dari tiga halaman utama, yaitu halaman daftar film, halaman detail film, dan halaman hasil rekomendasi. Ketiga halaman tersebut dirancang agar pengguna dapat melakukan proses rekomendasi secara bertahap, mulai dari melihat daftar film, memilih salah satu film, melihat informasi detail film, hingga memperoleh daftar film yang memiliki kemiripan tertinggi [6].

Halaman daftar film dirancang untuk menampilkan seluruh data film yang tersedia pada sistem. Pada halaman ini, pengguna dapat melihat informasi dasar film seperti judul, tahun rilis, genre, dan rating sebagai informasi tambahan. Selain itu, halaman daftar film juga dilengkapi dengan fitur pencarian dan filter genre agar pengguna dapat menemukan film yang diinginkan dengan lebih mudah.

Halaman detail film dirancang untuk menampilkan informasi lengkap mengenai film

yang dipilih oleh pengguna. Informasi yang ditampilkan meliputi judul film, tahun, genre, durasi, sutradara, pemain utama, rating, sinopsis, dan kata kunci. Pada halaman ini terdapat tombol “Rekomendasikan Film Mirip” yang digunakan sebagai input awal untuk menjalankan proses rekomendasi. Atribut utama yang digunakan dalam proses perhitungan rekomendasi adalah genre, sinopsis, kata kunci, sutradara, dan pemain utama, sedangkan informasi seperti rating, durasi, dan trailer hanya digunakan sebagai pelengkap tampilan antarmuka [3].

Halaman hasil rekomendasi dirancang untuk menampilkan daftar film yang memiliki nilai kemiripan tertinggi dengan film yang dipilih oleh pengguna. Pada halaman ini, sistem menampilkan film yang dipilih sebagai acuan rekomendasi, kemudian menampilkan beberapa film lain yang memiliki kemiripan berdasarkan hasil perhitungan menggunakan metode Content-Based Filtering, TF-IDF, dan Cosine Similarity [8]. Setiap film rekomendasi ditampilkan bersama nilai similarity agar pengguna dapat mengetahui tingkat kemiripan antara film yang dipilih dengan film hasil rekomendasi.

Dengan adanya perancangan antarmuka tersebut, sistem rekomendasi film diharapkan dapat memberikan gambaran yang jelas mengenai alur interaksi antara pengguna dan sistem. Penggunaan tiga halaman utama tersebut mendukung proses rekomendasi secara bertahap, yaitu pengguna memilih film melalui halaman daftar film, melihat informasi lengkap pada halaman detail film, kemudian memperoleh rekomendasi film yang relevan pada halaman hasil rekomendasi. Selain itu, penggunaan lima variabel utama dalam proses rekomendasi diharapkan dapat menghasilkan rekomendasi yang lebih sesuai karena sistem tidak hanya mempertimbangkan isi cerita dan tema film, tetapi juga memperhatikan sutradara dan pemain utama sebagai metadata pendukung.

III. HASIL DAN PEMBAHASAN

Hasil dan pembahasan pada penelitian ini disusun berdasarkan tahapan penelitian yang telah dijelaskan pada bagian metodologi.

Tahapan tersebut meliputi pengumpulan data, preprocessing, perancangan Content-Based Filtering, dan perancangan antarmuka. Penyusunan hasil dan pembahasan berdasarkan tahapan penelitian bertujuan agar proses perancangan sistem rekomendasi film dapat dijelaskan secara runtut, mulai dari data yang digunakan hingga rancangan tampilan sistem.

A. Pengumpulan Data

Tahap pengumpulan data dilakukan dengan menyusun data film yang akan digunakan sebagai data item dalam sistem rekomendasi. Data yang digunakan pada penelitian ini berjumlah 20 film Indonesia dengan atribut judul film, tahun, genre, sinopsis singkat, kata kunci, sutradara, dan pemain utama. Dari atribut tersebut, lima variabel utama yang digunakan dalam proses rekomendasi adalah genre, sinopsis, kata kunci, sutradara, dan pemain utama.

Jumlah data sebanyak 20 film digunakan sebagai dataset awal untuk menggambarkan proses perancangan dan ilustrasi penerapan metode Content-Based Filtering. Dataset ini belum dimaksudkan untuk mewakili seluruh keragaman film Indonesia, karena jumlah film yang digunakan masih terbatas. Oleh karena itu, hasil rekomendasi pada penelitian ini lebih tepat dipahami sebagai contoh hasil perhitungan berdasarkan data yang tersedia, bukan sebagai hasil generalisasi untuk seluruh film Indonesia.

Keterbatasan jumlah data dapat memengaruhi hasil rekomendasi, karena sistem rekomendasi sangat bergantung pada variasi dan jumlah item yang digunakan. Pada dataset yang kecil, hasil rekomendasi berpotensi lebih dipengaruhi oleh kesamaan langsung seperti film sekuel, sutradara yang sama, atau pemain utama yang sama. Oleh karena itu, pengembangan selanjutnya perlu menggunakan jumlah data film yang lebih besar dan lebih beragam agar hasil rekomendasi menjadi lebih representatif.

Pemilihan lima variabel tersebut didasarkan pada kebutuhan sistem dalam mengenali karakteristik film [5]. Genre digunakan untuk mengetahui kategori film, sinopsis digunakan untuk menggambarkan isi cerita, kata kunci digunakan untuk mewakili tema utama film, sutradara digunakan sebagai metadata pembuat

film, sedangkan pemain utama digunakan untuk menunjukkan aktor atau tokoh yang berperan penting dalam film [7]. Data film yang digunakan dalam penelitian ini ditampilkan pada Tabel II.

TABEL II. DATA FILM YANG DIGUNAKAN

No	Judul Film	Tahun	Genre	Sinopsis Singkat	Kata Kunci	Sutradara	Pemain Utama
1	Laskar Pelangi	2008	Drama, Pendidikan	Anak-anak di sekolah sederhana berusaha meraih mimpi melalui pendidikan dan persahabatan.	sekolah, mimpi, persahabatan, perjuangan	Riri Riza	Cut Mini, Ikranagara, Lukman Sardi, Zulfani
2	Habibi & Ainun	2012	Drama, Biografi, Romanis	Perjalanan hidup tokoh bangsa bersama pasangan yang mendukung karier dan perjuangannya.	cinta, tokoh, keluarga, perjuangan	Faozan Rizal	Reza Rahadian, Bunga Citra Lestari
3	Ngeri-Ngeri Sedap	2022	Drama, Komedi, Keluarga	Keluarga menghadapi konflik antara orang tua dan anak dengan pilihan hidup berbeda.	keluarga, budaya, konflik, orang tua	Benedictus Djojogadung	Arswendy Bening Swara, Tika Panggabean, Boris Bokir, Gita Bhebhita
4	Dilan 1990	2018	Romanis, Drama, Remaja	Hubungan remaja sekolah berkembang melalui perhatian, kenangan, dan komunikasi sederhana.	remaja, cinta, sekolah, kenangan	Fajar Bustomi, Pidi Baiq	Iqbaal Ramadhan, Vanesha Prescilla
5	Dilan 1991	2019	Romanis, Drama, Remaja	Kisah cinta remaja mulai diuji oleh perbedaan sikap, emosi, dan keadaan.	cinta, konflik, remaja, hubungan	Fajar Bustomi, Pidi Baiq	Iqbaal Ramadhan, Vanesha Prescilla

No	Judul Film	Tahun	Genre	Sinopsis Singkat	Kata Kunci	Sutradara	Pemain Utama
6	Milea: Suara dari Dilan	2020	Romanis, Drama	Kisah masa lalu hubungan diceritakan dari sudut pandang berbeda dengan nuansa nostalgia.	nostalgia, cinta, masa lalu, perpisahan	Fajar Bustomi, Pidi Baiq	Iqbaal Ramadhan, Vanesha Prescilla
7	Ada Apa dengan Cinta? 2	2016	Romanis, Drama	Dua orang yang pernah dekat kembali bertemu dan mencoba memahami perasaan lama.	cinta, reuni, masa lalu, perasaan	Riri Riza	Dian Sastrowardoyo, Nicholas Saputra
8	Ayat-Ayat Cinta	2008	Drama, Religi, Romanis	Mahasiswa menghadapi ujian cinta, keyakinan, dan tanggung jawab dalam kehidupan sosialnya.	religi, cinta, ujian hidup, mahasiswa	Hanung Bramantyo	Fedi Nuril, Rianti Cartwright, Carissa Putri
9	Miracle in Cell No. 7	2022	Drama, Keluarga	Hubungan ayah dan anak menggambarkan kasih sayang, ketulusan, dan perjuangan.	ayah, anak, keluarga, kasih sayang	Hanung Bramantyo	Vino G. Bastian, Graciella Abigail
10	Bila Esok Ibu Tiada	2024	Drama, Keluarga	Anak-anak menyadari pentingnya peran ibu dalam kehidupan keluarga.	ibu, keluarga, kehilangan, kasih sayang	Rudi Soedjarwo	Christine Hakim, Adinia Wirasti, Fedi Nuril, Amanda Manopo
11	My Stupid Boss	2016	Komedi	Karyawan menghadapi atasan unik yang sering membuat suasana kerja menjadi kacau.	kantor, bos, kerja, komedi	Upi Avianto	Bunga Citra Lestari, Reza Rahadian
12	Warkop DKI Reborn: Jangkrik Boss! Part 1	2016	Komedi	Tiga sahabat menghadapi berbagai kejadian lucu saat menjalankan tugas.	sahabat, humor, tugas, kekacauan	Anggy Umbara	Abimana Aryasatya, Vino G. Bastian, Tora Sudiro

No	Judul Film	Tahun	Genre	Sinopsis Singkat	Kata Kunci	Sutradara	Pemain Utama
13	Warkop DKI Reborn: Jangkrik Boss! Part 2	2017	Komedi, Petualangan	Petualangan tiga tokoh berlanjut dalam situasi lucu dan penuh kesalahpahaman.	petualangan, komedi, sahabat, misi	Anggy Umbara	Abimana Aryasatya, Vino G. Bastian, Tora Sudiro
14	Agak Laen	2024	Komedi, Drama	Sekelompok sahabat memperbaiki usaha hiburan, tetapi terlibat masalah semakin rumit.	persahabatan, usaha, humor, masalah	Muhadly Achor	Benedict Diona, Oki Rengga, Indra Jegel, Boris Bokir
15	Kang Makrom dan Pee Mak	2024	Komedi, Romantis	Cerita komedi romantis dengan unsur keluarga, persahabatan, dan kejadian tidak biasa.	komedi, cinta, keluarga, sahabat	Herwin Novianto	Vino G. Bastian, Marsha Timothy, Indro Warkop
16	KKN di Desa Penari	2022	Horor, Misteri	Mahasiswa mengalami kejadian aneh saat menjalankan kegiatan di desa terpencil.	desa, mahasiswa, misteri, horor	Awi Suryadi	Tissa Biani, Adinda Thomas, Achmad Megantara, Aulia Sarah
17	Pengabdian Setan	2017	Horor, Misteri	Keluarga mengalami gangguan setelah kehilangan anggota keluarga dan menemukan rahasia masa lalu.	keluarga, misteri, rumah, rahasia	Joko Anwar	Tara Basro, Bront Palarac, Endy Arfian
18	Pengabdian Setan 2: Community	2022	Horor, Misteri	Keluarga mencoba memulai hidup baru tetapi kembali menghadapi kejadian mengganggu.	keluarga, apartemen, misteri, masa lalu	Joko Anwar	Tara Basro, Endy Arfian, Nasar Annuz

No	Judul Film	Tahun	Genre	Sinopsis Singkat	Kata Kunci	Sutradara	Pemain Utama
19	Sewu Dino	2023	Horor, Misteri	Seorang perempuan menerima pekerjaan yang membawanya pada situasi penuh rahasia dan ancaman.	pekerjaan, rahasia, misteri, horor	Kimo Stamboel	Mikha Tambayon, Rio Dewanto, Givina Lukita
20	Badarawu di Desa Penari	2024	Horor, Misteri	Cerita berlatar desa dengan unsur legenda, aturan adat, dan kejadian misterius.	desa, legenda, misteri, adat	Kimo Stamboel	Aulia Sarah, Maudy Effrosina, Jourdy Pranata

Berdasarkan Tabel II, data film yang digunakan telah memenuhi kebutuhan minimal data item dalam proyek sistem rekomendasi. Data tersebut digunakan sebagai dasar dalam proses rekomendasi berbasis konten [4]. Semakin lengkap atribut film yang digunakan, maka semakin baik pula sistem dalam mengenali kemiripan antarfilm berdasarkan konten dan metadata [3].

B. Preprocessing

Tahap preprocessing dilakukan untuk menyiapkan data teks agar dapat diproses pada tahap perhitungan rekomendasi [9]. Pada tahap ini, lima variabel utama, yaitu genre, sinopsis, kata kunci, sutradara, dan pemain utama digabungkan menjadi satu dokumen teks untuk setiap film. Penggabungan variabel dilakukan agar setiap film memiliki representasi teks yang lebih lengkap [10].

Sebagai contoh, film “Dilan 1990” memiliki atribut genre romantis, drama, remaja; sinopsis yang menggambarkan kisah Dilan dan Milea; kata kunci cinta, remaja, sekolah, kenangan, dan hubungan; sutradara Fajar Bustomi dan Pidi Baiq; serta pemain utama Iqbaal Ramadhan dan Vanesha Prescilla. Seluruh atribut tersebut digabungkan menjadi satu dokumen teks yang mewakili karakteristik film “Dilan 1990”.

Setelah proses penggabungan variabel, dilakukan preprocessing teks yang meliputi case folding, tokenizing, stopword removal, dan

stemming [9]. Case folding digunakan untuk mengubah seluruh huruf menjadi huruf kecil [9]. Tokenizing digunakan untuk memecah kalimat menjadi kata-kata [9]. Stopword removal digunakan untuk menghapus kata umum yang tidak memiliki pengaruh besar terhadap makna rekomendasi [9]. Stemming digunakan untuk mengubah kata berimbuhan menjadi bentuk kata dasar [9]. Contoh tahapan preprocessing ditampilkan pada Tabel III.

Tabel III. Contoh Tahapan Preprocessing

No	Tahap Preprocessing	Hasil
1	Data Awal	Romantis, Drama, Remaja. Kisah Dilan dan Milea di masa sekolah yang menggambarkan cinta, kenangan, dan hubungan remaja. Fajar Bustomi, Pidi Baiq, Iqbaal Ramadhan, Vanesha Prescilla.
2	Case Folding	romantis, drama, remaja. kisah dilan dan milea di masa sekolah yang menggambarkan cinta, kenangan, dan hubungan remaja. fajar bustomi, pidi baiq, iqbaal ramadhan, vanesha prescilla.
3	Tokenizing	romantis, drama, remaja, kisah, dilan, milea, masa, sekolah, menggambarkan, cinta, kenangan, hubungan, remaja, fajar, bustomi, pidi, baiq, iqbaal, ramadhan, vanesha, prescilla
4	Stopword Removal	romantis, drama, remaja, kisah, dilan, milea, sekolah, cinta, kenangan, hubungan, remaja, fajar, bustomi, pidi, baiq, iqbaal, ramadhan, vanesha, prescilla
5	Stemming	romantis, drama, remaja, kisah, dilan, milea, sekolah, cinta, kenang, hubung, remaja, fajar, bustomi, pidi, baiq, iqbaal, ramadhan, vanesha, prescilla

Berdasarkan Tabel III, tahap preprocessing menghasilkan data teks yang lebih sederhana dan terstruktur. Data awal yang masih berbentuk gabungan kalimat dari beberapa atribut film diubah menjadi teks huruf kecil, kemudian dipisahkan menjadi token atau kata [9]. Selanjutnya, kata-kata umum yang tidak terlalu berpengaruh terhadap proses rekomendasi dihapus, lalu kata berimbuhan diubah menjadi bentuk dasar [9]. Hasil akhir preprocessing digunakan sebagai input pada tahap pembobotan TF-IDF [8].

C. Perancangan Content-Based Filtering

Tahap perancangan Content-Based Filtering dilakukan untuk menentukan proses rekomendasi film berdasarkan kemiripan konten dan metadata film [4]. Pada tahap ini, setiap film dianggap

sebagai satu dokumen, sedangkan gabungan genre, sinopsis, kata kunci, sutradara, dan pemain utama digunakan sebagai isi dokumen. Setelah data melalui tahap preprocessing, sistem melakukan pembobotan kata menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) [10].

TF-IDF digunakan untuk memberikan bobot pada setiap kata atau fitur yang terdapat pada dokumen film. Kata yang lebih sering muncul pada suatu film, tetapi tidak banyak muncul pada film lain, akan memiliki bobot yang lebih tinggi. Sebaliknya, kata yang terlalu umum dan muncul pada banyak film akan memiliki bobot yang lebih rendah. Setelah proses pembobotan TF-IDF, vektor setiap dokumen dinormalisasi menggunakan normalisasi L2 agar panjang vektor berada pada skala yang sama. Normalisasi ini dilakukan untuk mengurangi pengaruh perbedaan panjang teks antarfilm, karena setiap film dapat memiliki jumlah kata yang berbeda pada sinopsis, kata kunci, sutradara, dan pemain utama. Contoh hasil pembobotan TF-IDF ditampilkan pada Tabel IV.

TABEL IV. CONTOH HASIL PEMBOBOTAN TF-IDF

No	Film Acuan	Term/Fitur	DF	IDF	Bobot TF-IDF	Keterangan
1	Dilan 1990	remaja	2	2,946	2,946	Mewakili tema remaja pada film
2	Dilan 1990	sekolah	2	2,946	2,946	Mewakili latar cerita sekolah
3	Dilan 1990	fajar_bustomi	3	2,658	2,658	Mewakili variabel sutradara
4	Dilan 1990	pidi_baiq	3	2,658	2,658	Mewakili variabel sutradara
5	Dilan 1990	iqbaal_ramadhan	3	2,658	2,658	Mewakili variabel pemain utama
6	Dilan 1990	vanesha_prescilla	3	2,658	2,658	Mewakili variabel pemain utama
7	Dilan 1990	romantis	7	1,965	1,965	Mewakili genre film
8	KKN di Desa Penari	desa	2	2,946	2,946	Mewakili latar dan tema cerita
9	KKN di Desa Penari	mahasiswa	2	2,946	2,946	Mewakili tokoh dalam cerita
10	KKN di Desa Penari	horor	5	2,253	2,253	Mewakili genre film
11	KKN di Desa Penari	misteri	5	2,253	2,253	Mewakili genre dan suasana film
12	KKN di Desa Penari	awi_suryadi	1	3,351	3,351	Mewakili variabel sutradara

Berdasarkan Tabel IV, term atau fitur yang memiliki nilai TF-IDF tinggi menunjukkan karakteristik yang kuat dari suatu film. Pada film “Dilan 1990”, term seperti remaja, sekolah, fajar_bustomi, pidi_baiq, iqbaal_ramadhan, dan vanesha_prescilla memiliki pengaruh penting karena berkaitan langsung dengan genre, latar cerita, sutradara, dan pemain utama. Sementara itu, pada film “KKN di Desa Penari”, term seperti desa, mahasiswa, horor, misteri, dan awi_suryadi menjadi fitur penting yang menggambarkan karakteristik film horor-misteri tersebut.

Setelah pembobotan TF-IDF dilakukan, setiap film memiliki representasi dalam bentuk vector [12]. Vektor film acuan kemudian dibandingkan dengan vektor film lainnya menggunakan Cosine Similarity [11]. Perhitungan ini digunakan untuk mengetahui tingkat kemiripan antarfilm berdasarkan lima variabel utama. Nilai similarity berada pada rentang 0 sampai 1 [11]. Nilai yang semakin mendekati 1 menunjukkan bahwa film memiliki tingkat kemiripan yang semakin tinggi [11].

Contoh proses perhitungan Cosine Similarity untuk film acuan “Dilan 1990” ditampilkan pada Tabel V.

TABEL V. CONTOH PERHITUNGAN COSINE SIMILARITY FILM DILAN 1990

Peringkat	Film Acuan	Film Pembandingan	Dot Product Vektor Ternormalisasi	$\ A\ $ Setelah Normalisasi L2	$\ B\ $ Setelah Normalisasi L2	Perhitungan	Cosine Similarity
1	Dilan 1990	Dilan 1991	0,568	1	1	$0,568 / (1 \times 1)$	0,568
2	Dilan 1990	Milea : Suara dari Dilan	0,270	1	1	$0,270 / (1 \times 1)$	0,270
3	Dilan 1990	Laskar Pelangi	0,184	1	1	$0,184 / (1 \times 1)$	0,184
4	Dilan 1990	Kang Makrom Pec Mak	0,075	1	1	$0,075 / (1 \times 1)$	0,075
5	Dilan 1990	Habibie & Ainun	0,064	1	1	$0,064 / (1 \times 1)$	0,064

Berdasarkan Tabel V, film “Dilan 1991” memperoleh nilai similarity tertinggi terhadap “Dilan 1990”, yaitu 0,568. Setelah vektor TF-IDF dinormalisasi, tingkat kemiripan antarfilm dihitung menggunakan Cosine Similarity. Pada vektor yang telah dinormalisasi L2, panjang vektor bernilai 1, sehingga nilai Cosine Similarity setara dengan hasil dot product antarvektor ternormalisasi. Dengan demikian, nilai $(\|A\|)$ dan $(\|B\|)$ yang bernilai 1 pada tabel perhitungan bukan merupakan panjang vektor TF-IDF mentah, melainkan hasil dari proses normalisasi L2. Rumus yang digunakan adalah sebagai berikut:

Cosine:

$$(A, B) = (A \cdot B) / (\|A\| \|B\|)$$

Keterangan:

A adalah vektor TF-IDF film acuan, B adalah vektor TF-IDF film pembandingan, $A \cdot B$ adalah dot product kedua vektor, $\|A\|$ adalah panjang vektor film acuan, dan $\|B\|$ adalah panjang vektor film pembandingan.

Contoh proses perhitungan Cosine Similarity berikutnya dilakukan pada film acuan “KKN di Desa Penari”. Perhitungan ini bertujuan untuk melihat kemiripan film horor-misteri dengan film lain yang memiliki karakteristik serupa. Hasil perhitungan ditampilkan pada Tabel VI.

TABEL VI. CONTOH PERHITUNGAN COSINE SIMILARITY FILM KKN DI DESA PENARI

Peringkat	Film Acuan	Film Pembandingan	Dot Product Vektor Ternormalisasi	$\ A\ $ Setelah Normalisasi L2	$\ B\ $ Setelah Normalisasi L2	Perhitungan	Cosine Similarity
1	KKN di Desa Penari	Badarawudi di Desa Penari	0,304	1	1	$0,304 / (1 \times 1)$	0,304
2	KKN di Desa Penari	Pengabdian Setan	0,157	1	1	$0,157 / (1 \times 1)$	0,157

3	KKN di Desa Penari	Pengabdian Setan 2: Communion	0,156	1	1	$0,156 / (1 \times 1)$	0,156
4	KKN di Desa Penari	Sewu Dino	0,133	1	1	$0,133 / (1 \times 1)$	0,133
5	KKN di Desa Penari	Ayat-Ayat Cinta	0,108	1	1	$0,108 / (1 \times 1)$	0,108

Berdasarkan Tabel VI, film “Badarawuhi di Desa Penari” memperoleh nilai similarity tertinggi terhadap “KKN di Desa Penari”, yaitu 0,304. Kemiripan ini dipengaruhi oleh kesamaan tema desa, horor, misteri, dan keterkaitan tokoh. Film lain seperti “Pengabdian Setan”, “Pengabdian Setan 2: Communion”, dan “Sewu Dino” juga memperoleh nilai similarity karena memiliki genre dan suasana cerita yang berdekatan.

Keterangan: Nilai $(||A||)$ dan $(||B||)$ bernilai 1 karena vektor TF-IDF telah melalui proses normalisasi L2. Oleh karena itu, nilai similarity pada tabel diperoleh dari dot product antarvektor yang telah dinormalisasi. Nilai tersebut digunakan sebagai contoh perhitungan berdasarkan dataset terbatas yang digunakan dalam penelitian.

Contoh hasil pembobotan TF-IDF ditampilkan pada Tabel IV. Selanjutnya, contoh perhitungan Cosine Similarity untuk film acuan “Dilan 1990” ditampilkan pada Tabel V, sedangkan contoh perhitungan Cosine Similarity untuk film acuan “KKN di Desa Penari” ditampilkan pada Tabel VI. Perhitungan tersebut digunakan sebagai ilustrasi proses kerja metode, bukan sebagai evaluasi akhir terhadap efektivitas sistem rekomendasi.

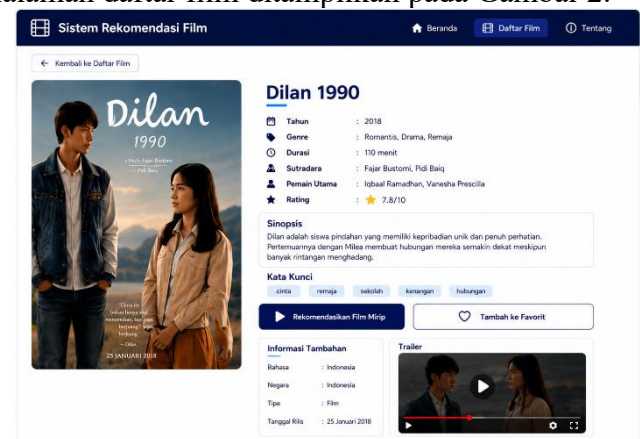
Setelah nilai Cosine Similarity diperoleh, sistem mengurutkan film berdasarkan nilai

similarity tertinggi [13]. Film dengan nilai similarity terbesar akan ditempatkan pada urutan teratas sebagai rekomendasi utama [13]. Hasil ini menunjukkan bahwa metode Content-Based Filtering dapat digunakan untuk memberikan rekomendasi film berdasarkan kesamaan karakteristik konten dan metadata [7]. Film yang memiliki kesamaan genre, kata kunci, sutradara, dan pemain utama cenderung memperoleh nilai similarity yang lebih tinggi.

D. Perancangan Antarmuka

Tahap perancangan antarmuka dilakukan untuk memberikan gambaran tampilan sistem rekomendasi film yang akan digunakan oleh pengguna. Antarmuka yang dirancang terdiri dari tiga halaman utama, yaitu halaman daftar film, halaman detail film, dan halaman hasil rekomendasi. Ketiga halaman tersebut dibuat untuk mendukung alur interaksi pengguna, mulai dari melihat daftar film, memilih salah satu film, melihat detail film, hingga memperoleh daftar rekomendasi film [6].

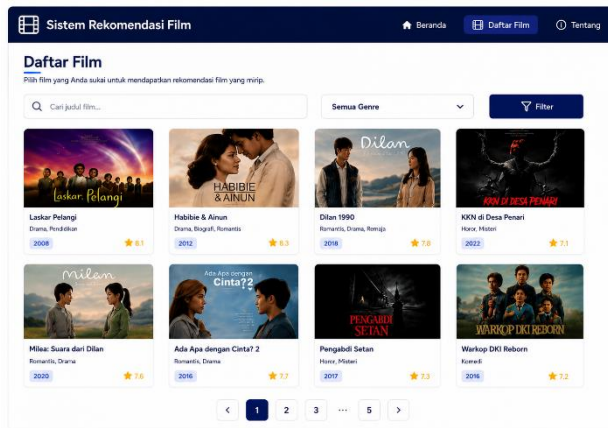
Halaman daftar film digunakan untuk menampilkan seluruh data film yang tersedia pada sistem. Pada halaman ini, pengguna dapat melihat informasi dasar film seperti judul, tahun rilis, genre, dan rating sebagai informasi tambahan. Halaman daftar film juga dilengkapi dengan fitur pencarian dan filter genre agar pengguna dapat menemukan film yang diinginkan dengan lebih mudah [14]. Rancangan halaman daftar film ditampilkan pada Gambar 2.



Gambar 3. Rancangan Halaman Detail Film

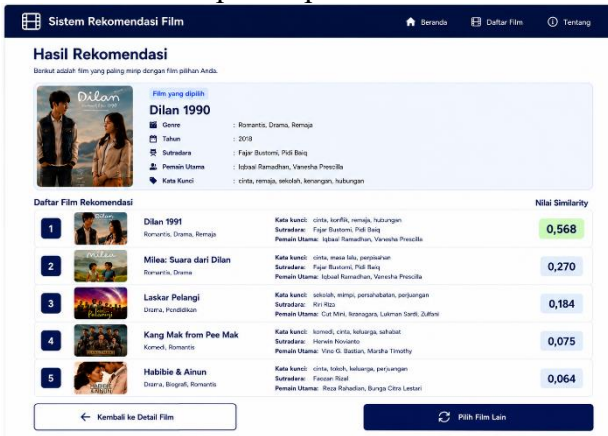
Halaman detail film digunakan untuk menampilkan informasi lengkap mengenai film yang dipilih oleh pengguna. Informasi yang

ditampilkan meliputi judul film, tahun, genre, durasi, sutradara, pemain utama, rating, sinopsis, dan kata kunci [3]. Pada halaman ini terdapat tombol “Rekomendasikan Film Mirip” yang digunakan sebagai input awal untuk menjalankan proses rekomendasi [6]. Rancangan halaman detail film ditampilkan pada Gambar 3.



Gambar 2. Rancangan Halaman Daftar Film

Halaman hasil rekomendasi digunakan untuk menampilkan daftar film yang memiliki nilai similarity tertinggi dengan film yang dipilih oleh pengguna [13]. Pada halaman ini, sistem menampilkan film acuan, daftar film rekomendasi, dan nilai similarity dari setiap film rekomendasi [12]. Rancangan halaman hasil rekomendasi ditampilkan pada Gambar 4.



Gambar 4. Rancangan Halaman Hasil Rekomendasi

Berdasarkan hasil perancangan antarmuka, sistem rekomendasi film memiliki alur penggunaan yang sederhana dan mudah dipahami. Pengguna dapat memilih film melalui halaman daftar film, melihat informasi lengkap pada halaman detail film, kemudian memperoleh rekomendasi film yang relevan pada halaman hasil rekomendasi [6]. Perancangan antarmuka

ini mendukung proses rekomendasi agar hasil perhitungan dapat ditampilkan secara jelas kepada pengguna.

E. Pembahasan

Berdasarkan hasil perancangan, metode Content-Based Filtering sesuai digunakan pada sistem rekomendasi film karena metode ini dapat memanfaatkan karakteristik item sebagai dasar rekomendasi [4]. Penggunaan lima variabel utama, yaitu genre, sinopsis, kata kunci, sutradara, dan pemain utama, membuat sistem dapat mengenali karakteristik film secara lebih lengkap [5]. Sistem tidak hanya membandingkan film berdasarkan isi cerita dan tema, tetapi juga memperhatikan metadata film seperti sutradara dan pemain utama [7].

Penggunaan TF-IDF membantu sistem memberikan bobot pada kata yang memiliki nilai penting dalam dokumen teks film [8]. Berdasarkan contoh hasil pembobotan, term yang lebih khas seperti nama sutradara, pemain utama, tema cerita, dan kata kunci tertentu dapat membantu membedakan satu film dengan film lainnya [10]. Selanjutnya, Cosine Similarity digunakan untuk menghitung kedekatan antarfilm berdasarkan hasil pembobotan tersebut [11]. Hasil contoh rekomendasi menunjukkan bahwa film yang memiliki kesamaan kuat pada genre, kata kunci, sutradara, dan pemain utama memperoleh nilai similarity yang lebih tinggi [12].

Meskipun rancangan sistem telah mampu menggambarkan proses rekomendasi berdasarkan lima variabel, penelitian ini masih memiliki keterbatasan. Jumlah data yang digunakan masih terbatas, yaitu 20 data film. Selain itu, kualitas hasil rekomendasi sangat bergantung pada kelengkapan dan ketepatan atribut film yang digunakan [15]. Jika sinopsis terlalu singkat, kata kunci terlalu umum, atau data sutradara dan pemain utama tidak lengkap, maka hasil rekomendasi dapat menjadi kurang relevan [9]. Oleh karena itu, pengembangan selanjutnya dapat dilakukan dengan menambah jumlah data film, memperbaiki kualitas atribut, serta mengimplementasikan rancangan ini ke dalam aplikasi berbasis web [1].

F. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan. Pertama, jumlah data film yang digunakan masih terbatas, yaitu 20 data film. Jumlah tersebut digunakan untuk memenuhi kebutuhan data awal dalam perancangan dan ilustrasi metode, tetapi belum cukup untuk mewakili keragaman film Indonesia secara luas. Dengan jumlah data yang terbatas, hasil rekomendasi dapat dipengaruhi oleh kesamaan langsung seperti sekuel, sutradara yang sama, atau pemain utama yang sama.

Kedua, penelitian ini masih berfokus pada tahap perancangan sistem, meliputi perancangan data, preprocessing, metode Content-Based Filtering, perhitungan TF-IDF dan Cosine Similarity, serta perancangan antarmuka. Penelitian ini belum sampai pada tahap implementasi sistem secara penuh dan belum melakukan evaluasi langsung kepada pengguna. Oleh karena itu, efektivitas sistem dalam membantu pengguna memilih film belum dapat dibuktikan secara empiris.

Ketiga, hasil perhitungan yang ditampilkan pada penelitian ini digunakan sebagai contoh ilustrasi proses kerja metode Content-Based Filtering. Untuk menghasilkan sistem rekomendasi yang lebih akurat dan representatif, penelitian selanjutnya perlu menggunakan dataset yang lebih besar, atribut film yang lebih lengkap, implementasi sistem secara nyata, serta evaluasi pengguna untuk mengukur kualitas rekomendasi yang dihasilkan.

IV. KESIMPULAN

Berdasarkan hasil perancangan yang telah dilakukan, sistem rekomendasi film dapat dirancang menggunakan metode Content-Based Filtering dengan memanfaatkan lima variabel utama, yaitu genre, sinopsis, kata kunci, sutradara, dan pemain utama. Kelima variabel tersebut digabungkan menjadi dokumen teks film, kemudian diproses melalui tahapan preprocessing, pembobotan TF-IDF, normalisasi vektor, dan perhitungan Cosine Similarity. Hasil perhitungan menunjukkan bahwa pada film acuan “Dilan 1990”, film yang direkomendasikan sebagai rekomendasi utama adalah “Dilan 1991” dengan nilai similarity sebesar 0,568. Sementara

itu, pada film acuan “KKN di Desa Penari”, film yang direkomendasikan sebagai rekomendasi utama adalah “Badarawuhi di Desa Penari” dengan nilai similarity sebesar 0,304.

Hasil tersebut menunjukkan bahwa metode Content-Based Filtering dapat digunakan untuk menggambarkan proses rekomendasi film berdasarkan kesamaan konten dan metadata. Namun, hasil penelitian ini masih bersifat ilustratif karena dataset yang digunakan hanya berjumlah 20 film dan penelitian masih terbatas pada tahap perancangan. Dengan demikian, penelitian ini belum dapat membuktikan efektivitas sistem secara empiris dalam membantu pengguna memilih film. Rancangan yang dihasilkan dapat dijadikan dasar untuk pengembangan aplikasi rekomendasi film berbasis web pada penelitian selanjutnya. Pengembangan berikutnya perlu menggunakan dataset yang lebih besar dan beragam, mengimplementasikan sistem secara penuh, serta melakukan evaluasi pengguna untuk mengetahui kualitas rekomendasi yang dihasilkan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pengampu mata kuliah Sistem Rekomendasi yang telah memberikan arahan dalam penyusunan project ini. Penulis juga mengucapkan terima kasih kepada seluruh anggota kelompok yang telah berkontribusi dalam pengumpulan data dan penyusunan rancangan sistem.

REFERENSI

- [1] P. Lestari, “Sistem Rekomendasi untuk Maksimalisasi Industri Film dengan Metode Demographic Filtering dan Content Based Filtering,” *J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 1–10, 2024, doi: 10.54082/jiki.104.
- [2] T. Ridwansyah, B. Subartini, and S. Sylviani, “Penerapan Metode Content-Based Filtering pada Sistem Rekomendasi,” vol. 4, no. 2, 2024.
- [3] R. Eka, P. Wardhani, and A. N. Rohman, “SISTEM REKOMENDASI FILM BERBASIS CONTENT-BASED FILTERING,” vol. 11, no. 1, pp. 1809–1817, 2026.
- [4] M. Fajriansyah, P. P. Adikara, and A. W. Widodo, “Sistem Rekomendasi Film Menggunakan Content Based Filtering,” vol. 5, no. 6, 2021.
- [5] C. Y. Fuady, V. C. Mawardi, and T. Sutrisno, “IMPLEMENTASI METODE CONTENT BASED FILTERING UNTUK REKOMENDASI FILM BERDASARKAN GENRE DAN SINOPSIS,” vol. 01, no. 02, pp. 227–236, 2023.

- [6] V. No, D. T. Santoso, V. Atina, D. Hartanti, and V. No, "Infotek : Jurnal Informatika dan Teknologi Prototipe Sistem Rekomendasi Film Indonesia Menggunakan Pendekatan Content Based Filtering dan Metode Vector Space Model Infotek : Jurnal Informatika dan Teknologi Film , menurut Pratista dalam bukunya " Memahami," vol. 7, no. 2, 2024.
- [7] A. Zakharia, A. D. Ulhaq, A. B. Suryono, and N. C. Nugroho, "Sistem Rekomendasi Film Indonesia Menggunakan Metode Content-Based Filtering," vol. 2, no. 4, pp. 671–678, 2024.
- [8] A. Febrian and E. D. Permana, "Sistem Rekomendasi Film Menggunakan Metode Content Based Filtering Dengan Algoritma Tf-Idf," *Al-Aqlu J. Mat. Tek. dan Sains*, vol. 4, no. 1, pp. 19–25, 2026, doi: 10.59896/aqlu.v4i1.494.
- [9] R. M. Harahap and A. N. Rachman, "Film Kemiripan Deskripsi Cerita Menggunakan," vol. 13, no. 3, 2025.
- [10] A. H. J. P. J. P. A. T. Wibowo, "Movie Recommendation System Based on Synopsis Using Content-Based Filtering with TF-IDF and Cosine Similarity," *Int. J. Inf. Commun. Technol.*, no. Vol. 9 No. 2 (2023): Vol.9 No. 2 Dec 2023, pp. 1–14, 2023.
- [11] T. Wahyu Intan Permadani, D. Adi Prasetya, E. Daniati, and P. Korespondens, "Sistem Rekomendasi Film Berdasarkan Genre Menggunakan Metode Content-Based Filtering dengan Algoritma Cosine Similarity," *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.*, vol. 9, no. 3, pp. 1835–1843, 2025, [Online]. Available: <https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/7189>
- [12] E. Z. Irabela, "Implementation of Content-Based Cosine Similarity Algorithm With TF-IDF and SBERT for Movie Recommendation," vol. 09, no. 01, pp. 114–122, 2026.
- [13] B. Syava, A. Bana, T. Informatika, F. Teknik, U. Nusantara, and P. Kediri, "Sistem Rekomendasi Film Berdasarkan Deskripsi Pengguna," vol. 5, pp. 97–105, 2026.
- [14] J. Ring, R. Utara, and D. I. Yogyakarta, "Penerapan Content-Based Filtering untuk Rekomendasi Film Netflix Berdasarkan Genre dan Klasifikasi Usia Penonton," vol. 5, no. November, 2025.
- [15] S. Lestari and M. M. Ramdhani, "Manajemen Informatika dan Komunikasi SISTEM REKOMENDASI FILM MENGGUNAKAN METODE CONTENT-BASED FILTERING STUDI KASUS MATERI DATA MINING DI SMK IDN BOARDING Jurnal Indonesia : Manajemen Informatika dan Komunikasi," vol. 4, no. 3, pp. 1581–1587, 2023.