

Prototipe Sistem Rekomendasi Film Indonesia Menggunakan Pendekatan Content Based Filtering dan Metode Vector Space Model

Daniel Theo Santoso^{1*}, Vihi Atina², Dwi Hartanti³

^{1,3}Program Studi Teknik Informatika, Universitas Duta Bangsa

²Program Studi Manajemen Informatika, Universitas Duta Bangsa

*202030298@mhs.udb.ac.id

Abstrak

Film merupakan perpaduan antara narasi dan aspek sinematografi dalam bentuk audio-visual. Film juga memberikan pengalaman visual, auditif, dan emosional yang menarik. Industri film di Indonesia menunjukkan pertumbuhan signifikan. Pada tahun 2015, penonton bioskop berjumlah 16,2 juta dan meningkat menjadi 51,2 juta pada tahun 2019. Pandemi COVID-19 mengakibatkan penurunan drastis pada jumlah penonton, dengan hanya 4,5 juta pada tahun 2021. Namun, industri ini berhasil pulih dengan cepat, mencapai 54,07 juta penonton pada tahun 2022 dan 55 juta pada tahun 2023, dan diperkirakan akan mencapai 60 juta pada tahun 2024. Dengan meningkatnya ketersediaan film dan platform streaming, Menemukan film yang sesuai dengan preferensi sering menjadi tantangan bagi pengguna. Penelitian ini mengusulkan sistem rekomendasi film menggunakan metode Content Based Filtering dan Vector Space Model agar pengguna dapat lebih mudah menemukan film yang sesuai dengan preferensi mereka. Content Based Filtering merekomendasikan film berdasarkan kesamaan konten, sedangkan Vector Space Model mengukur kesamaan konten menggunakan metrik seperti cosine similarity atau euclidean distance. Berdasarkan percobaan menggunakan 20 sampel film, hasil menunjukkan bahwa sistem yang dirancang dapat memberikan rekomendasi film secara akurat dan tepat berdasarkan kata kunci yang diberikan oleh pengguna. Film "Pengabdian Setan 2: Communion" memiliki nilai similarity tertinggi, yaitu 0.3018, diikuti oleh "Nana", "Menjelang Magrib", "Jailangkung: Sandekala", dan "Qorin" dengan nilai similarity masing-masing 0.0865, 0.0138, 0.0136, dan 0.0125.

Kata kunci: Sistem Rekomendasi, Film, Content Based Filtering, Vector Space Model

Abstract

Film represents a combination of narrative and cinematographic aspects in an audio-visual form. Films also provide an engaging visual, auditory, and emotional experience. The film industry in Indonesia has shown significant growth. In 2015, cinema audiences numbered 16.2 million, increasing to 51.2 million in 2019. The COVID-19 pandemic caused a drastic decline in the number of viewers, with only 4.5 million in 2021. However, the industry quickly recovered, reaching 54.07 million viewers in 2022 and 55 million in 2023, and is projected to reach 60 million in 2024. With the increasing availability of films and streaming platforms, finding movies that match user preferences often becomes a challenge. This study proposes a film recommendation system using the Content-Based Filtering method and the Vector Space Model to help users more easily find movies that match their preferences. Content-Based Filtering recommends films based on content similarity, while the Vector Space Model measures content similarity using metrics such as cosine similarity or euclidean distance. Based on experiments using 20 sample films, the results show that the designed system can provide accurate and precise movie recommendations based on keywords provided by users. The film "Pengabdian Setan 2: Communion" has the highest similarity score of 0.3018, followed by "Nana," "Menjelang Magrib," "Jailangkung: Sandekala," and "Qorin," with similarity scores of 0.0865, 0.0138, 0.0136, and 0.0125, respectively.

Keywords: Recommendation System, Movie, Content Based Filtering, Vector Space Model.

1. Pendahuluan

Film, menurut Pratista dalam bukunya "Memahami Film", adalah media audio-visual yang menggabungkan unsur naratif dan sinematik [1]. Film menjadi salah satu media hiburan yang digemari karena memberikan pengalaman visual, auditif, dan emosional yang menarik. Banyak orang mengisi waktu luang mereka dengan menonton film di bioskop, televisi, atau platform online. Industri film di Indonesia telah berkembang pesat, dengan jumlah penonton bioskop meningkat dari 16,2 juta pada tahun 2015 menjadi 51,2 juta pada tahun 2019. Namun, pandemi COVID-19 mengakibatkan penurunan drastis jumlah penonton menjadi 19 juta pada tahun 2020 dan 4,5 juta pada tahun 2021. Meski demikian, industri ini pulih dengan cepat, mencapai 54,07 juta penonton pada tahun 2022 dan diperkirakan akan mencapai 60 juta pada tahun 2024 [2], [3], [4].

Film menawarkan berbagai pilihan sesuai selera penonton, dikategorikan berdasarkan genre seperti drama, komedi, aksi, horor, dan lainnya. Namun, dengan semakin banyaknya film yang tersedia, penonton sering kali kesulitan menemukan film yang sesuai preferensi. Laporan Nielsen, State of Play mencatat bahwa penonton di Amerika Serikat kini memiliki akses ke lebih dari 817.000 judul program, meningkat 18% dari 646.000 judul pada akhir tahun 2019 [5]. Di Indonesia, Klik Film menjadi repositori film

terbesar dengan koleksi 450 film nasional pada tahun 2020, diikuti Disney+ Hotstar dan Netflix dengan koleksi 220 dan 196 film Indonesia masing-masing [6]. Keberagaman ini, meski memberikan banyak pilihan, sering kali membuat penonton bingung dan kesulitan memilih film yang sesuai selera, menyebabkan rasa kecewa atau tidak puas.

Sebagai solusi, penelitian ini mengusulkan pembangunan sistem rekomendasi film menggunakan pendekatan Content-Based Filtering dan metode Vector Space Model. Content-Based Filtering merekomendasikan film berdasarkan kesamaan konten antara film yang disukai pengguna dengan film lain dalam database. Vector Space Model merepresentasikan konten film sebagai vektor dalam ruang multidimensi, memungkinkan pengukuran kesamaan konten film menggunakan metrik seperti cosine similarity atau euclidean distance. Dengan pendekatan ini, penelitian ini berharap dapat menghasilkan sistem rekomendasi yang memberikan saran film relevan, menarik, dan bervariasi bagi pengguna..

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Adapun penelitian sebelumnya yang digunakan sebagai acuan dalam penulisan sesuai dengan yang akan di teliti sebagai pendukung penelitian ini, antara lain:

- Berdasarkan penelitian berjudul "Sistem Rekomendasi Film Menggunakan Metode K-Means Clustering" yang ditulis oleh Herdita Mutiasari, Tito Waluyo Purboyo, dan Ratna Astuti Nugrahaeni (2021), penelitian ini bertujuan untuk mengembangkan sebuah aplikasi yang dapat membantu pengguna mendapatkan rekomendasi film sesuai dengan preferensi mereka. Hasil penelitian ini menunjukkan bahwa penggunaan algoritma K-Means dengan 4 bobot preferensi menghasilkan tingkat kesalahan yang lebih rendah, yaitu 0,44%, dibandingkan dengan penggunaan 3 atau 5 bobot preferensi [7].
- Penelitian oleh Dodi Nugraha, Tito Waluyo Purboyo, dan Ratna Astuti Nugrahaeni (2021) berjudul "Sistem Rekomendasi Film Menggunakan Metode User Based Collaborative Filtering" bertujuan mengembangkan sistem yang memberikan rekomendasi film untuk menarik minat penonton terhadap berbagai film. Sistem ini memberikan informasi jelas kepada masyarakat tentang film-film yang mungkin ingin mereka tonton. Rekomendasi dihasilkan menggunakan metode Machine Learning, dengan aplikasi yang menggunakan 5 bobot preferensi dalam algoritma User-Based Collaborative Filtering. Hasil penelitian menunjukkan bahwa aplikasi ini memiliki tingkat kesalahan yang lebih rendah, yaitu 0,73%, dibandingkan dengan penggunaan 3 atau 4 preferensi [8].
- Penelitian oleh Arif Akbarul Huda, Rohmad Fajarudin, dan Arifiyanto Hadinegoro (2022) berjudul "Sistem Rekomendasi Content-Based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita" membahas masalah yang muncul akibat keberagaman dan banyaknya artikel berita yang terus diproduksi. Meski jumlah artikel yang tersedia terus bertambah, mahasiswa sering kewalahan memilih artikel yang sesuai dengan minat mereka. Hasil evaluasi dengan nilai Recall@5 sekitar 73% dan Recall@10 sekitar 80% menunjukkan bahwa sistem rekomendasi yang dikembangkan mampu memberikan daftar rekomendasi artikel yang relevan. Dengan ini, mahasiswa dapat lebih mudah menemukan artikel yang sesuai dengan minat dan kebutuhan mereka [9].
- Penelitian oleh Joni, Andy, dan Karin Wibowo (2021) berjudul "Perancangan Website Rekomendasi Film dengan Metode User-Based Collaborative Filtering" membahas kebingungan dalam memilih film di tengah banyaknya pilihan. Sering kali, penonton kesulitan menentukan film yang ingin ditonton karena variasi kualitas dalam genre yang sama. Untuk mengatasi masalah ini, metode User-Based Collaborative Filtering diterapkan, memberikan rekomendasi

berdasarkan pola rating dan ulasan pengguna lain. Hasil penelitian menunjukkan bahwa sistem rekomendasi ini mampu memberikan saran film yang lebih akurat dan relevan, memudahkan penonton menemukan film berdasarkan preferensi mereka [10].

- Penelitian oleh Eggy Ryana Agustian, Munir, dan Eddy Prasetyo Nugroho (2020) berjudul "Sistem Rekomendasi Film Menggunakan Metode Collaborative Filtering dan K-Nearest Neighbors" membahas masalah pemilihan film di tengah banyaknya pilihan. Dengan pesatnya perkembangan industri film, banyak penonton kesulitan memilih film yang ingin ditonton. Penelitian ini menggunakan metode Collaborative Filtering (CF) untuk memberikan rekomendasi film berdasarkan preferensi pengguna lain, memprediksi rating film dari pola rating pengguna lain yang memiliki preferensi mirip. Aplikasi berbasis website ini memungkinkan pengguna melihat rekomendasi film yang sesuai dengan selera mereka. Metode weight sum digunakan untuk mengurutkan hasil rekomendasi dari rating tertinggi ke terendah, memudahkan penonton dalam memilih film [11].

2.2. Landasan Teori

1. Sistem Rekomendasi

Sistem rekomendasi dirancang sebagai model aplikasi yang memfasilitasi pengguna dalam

mengidentifikasi dan merekomendasikan produk yang paling sesuai dengan preferensi serta keinginan mereka, sekaligus memudahkan pengguna dalam pengambilan keputusan [12]. Metode-metode seperti Collaborative Filtering, Content-Based Filtering, dan Hybrid Filtering digunakan untuk membuat rekomendasi, masing-masing dengan kelebihan dan kekurangan tersendiri dalam hal teknik, dataset, platform, dan metrik kinerja yang digunakan.

Dengan adanya perubahan pola penggunaan komputer, personalisasi, dan kemudahan mengakses internet, informasi online menjadi sangat banyak dan beragam. Sistem rekomendasi adalah alat yang dapat membantu menyortir informasi online sesuai dengan kebutuhan dan minat pengguna [13].

2. Content Based Filtering

Content Based Filtering merupakan teknik rekomendasi yang menggunakan sifat, ciri, atau elemen dari suatu item, yang kemudian digunakan untuk mengukur kemiripan antara item-item dalam memberikan rekomendasi [14]. Teknik ini tidak membutuhkan informasi tentang preferensi pengguna lain, sehingga lebih mudah diimplementasikan dan dijelaskan.

3. Vector Space Model

Vector Space Model (VSM) adalah sebuah model yang digunakan untuk mengukur kesamaan atau relevansi antara dokumen dan query dalam pengambilan informasi. Model ini

merepresentasikan dokumen dan query sebagai vektor dari term yang ada dalam korpus, dengan menggunakan bobot term seperti TF-IDF untuk menunjukkan pentingnya serta spesifiknya term untuk dokumen atau query. Model ini juga menggunakan cosine similarity untuk menghitung kesamaan antara vektor dokumen dan query. Model ini memiliki keuntungan seperti mudah diimplementasikan, intuitif, dan fleksibel [15].

4. Film

Film adalah karya seni yang tercipta dari imajinasi dan kreativitas orang-orang yang berperan dalam proses pembuatannya. Film bisa berupa selaput tipis dari seluloid yang berisi gambar negatif atau positif, atau bisa juga berupa cerita dalam hidup. Film bisa menjadi media untuk mengekspresikan kemampuan kreatif yang bisa menciptakan realitas fiktif yang berbeda dari realitas nyata. Film bisa menampilkan keindahan atau hiburan bagi penontonnya. Film telah menjadi media yang semakin populer di kalangan masyarakat sebagai sarana untuk menghibur dan menghadirkan beragam cerita, musik, drama, komedi, dan berbagai aspek teknis lainnya [16]. Film bisa dibedakan berdasarkan genre atau kategori yang memiliki kesamaan dalam bentuk, tema, suasana dan sebagainya [17]

3. Metode Penelitian

3.1. Metode Pengumpulan Data

1. Observasi

Observasi adalah metode yang digunakan untuk melakukan pengamatan langsung terhadap objek penelitian, dalam hal ini adalah data film yang diambil dari situs web themoviedb.org. Tujuan dari observasi ini adalah untuk mengidentifikasi variabel-variabel yang terdapat dalam data tersebut. Dengan melakukan observasi, penelitian dapat memahami secara langsung berbagai aspek dan karakteristik yang terdapat pada setiap variabel dalam dataset film tersebut.

1. Studi Pustaka

Studi pustaka merupakan langkah yang dilakukan untuk mencari jurnal, laporan, artikel ilmiah, dan referensi lain yang berkaitan dengan penelitian. Tujuan utamanya adalah untuk mengumpulkan informasi yang relevan dan mendalam tentang topik penelitian dari berbagai sumber yang terpercaya..

3.2. Metode Pengembangan Sistem

Metode Prototipe adalah suatu pendekatan dalam pengembangan perangkat lunak di mana sebuah prototipe atau model awal dibuat, diuji, dan diperbaiki hingga mencapai prototipe yang dapat diterima. Prototipe adalah visualisasi sederhana dari produk untuk menguji konsep. Prototipe memungkinkan pembangunan versi awal dari produk untuk menguji konsep dan proses.

Keuntungan dari model Prototipe termasuk mendapatkan umpan balik pengguna di tahap awal proyek, mengidentifikasi fungsi yang hilang dengan mudah, dan mengurangi biaya. Berikut adalah langkah-langkah dalam metode Prototipe:

1. Definisi: Pengumpulan dan Analisis Persyaratan: Menentukan kebutuhan sistem melalui wawancara dengan pengguna.
2. Desain Cepat: Membuat desain dasar sistem untuk memberikan gambaran awal kepada pengguna.
3. Pembuatan Prototipe: Mengembangkan Prototipe awal berdasarkan desain cepat untuk memahami lebih lanjut tentang desain.
4. Evaluasi Awal oleh Pengguna: Menyajikan sistem kepada klien untuk pengujian awal dan mengumpulkan umpan balik.
5. Pengembangan Prototipe yang Diperbaiki: Memperbaiki prototipe berdasarkan umpan balik pengguna.
6. Implementasi Produk dan Pemeliharaan: Menguji sistem akhir sepenuhnya, mendistribusikannya ke produksi, dan melakukan pemeliharaan rutin..

4. Hasil dan Pembahasan

4.1. Pembuatan Prototipe

Penelitian ini menggunakan 20 sampel film Indonesia yang dirilis pada tahun 2022. Hasil perhitungan ditampilkan menggunakan satu sampel film berjudul "Pengabdian Setan 2:

Communion". Langkah pertama dalam perancangan sistem rekomendasi adalah pengumpulan data melalui API. Sebelum disimpan ke database, data tersebut diproses terlebih dahulu untuk mengekstraksi kata-kata penting dari judul, genre, dan ringkasan film. Setelah data disimpan, dilakukan perhitungan menggunakan metode Vector Space Model dengan Cosine Similarity untuk mendapatkan rekomendasi film berdasarkan kata kunci yang dimasukkan. Berikut adalah proses dan hasil dari tahap preprocessing dan perhitungan tersebut:

4.2. Text Preprocessing

Text Preprocessing merupakan langkah yang dilakukan untuk membersihkan dan mengolah teks agar siap untuk diproses pada tahap selanjutnya. Contoh data yang akan digunakan tercantum pada Tabel 1.

Tabel 1. Sampel Film

No.	Variabel	Teks
1	Judul	Pengabdian Setan 2: Communion
	Ringkasan	Setelah pindah dari rumah mereka ke gedung apartemen, teror baru menunggu keluarga Rini.
	Genre	Drama, Horor, Misteri
2	Judul	Sri Asih
	Ringkasan	Alana menemukan kebenaran tentang asalnya: dia bukan manusia biasa. Dia mungkin menjadi hadiah untuk kemanusiaan dan menjadi pelindungnya sebagai Sri Asih. Atau kehancuran, jika dia tidak bisa mengendalikan kemarahannya.
	Genre	Aksi, Petualangan, Fiksi Ilmiah, Fantasi, Drama

1. Cleaning

Cleaning adalah tahap di mana teks dibersihkan dari elemen-elemen yang tidak diinginkan atau tidak relevan untuk memastikan bahwa teks dalam kondisi optimal sebelum analisis lebih lanjut, seperti tanda baca, karakter khusus, dan spasi yang berlebihan. Hasil Cleaning dari data sampel dapat dilihat pada Tabel 2.

Tabel 2. Hasil Cleaning

Masukkan	Hasil
Pengabdi Setan 2: Communion. Setelah pindah dari rumah mereka ke gedung apartemen, teror baru menunggu keluarga Rini. Drama, Horor, Misteri	Pengabdi Setan Communion Setelah pindah dari rumah mereka ke gedung apartemen teror baru menunggu keluarga Rini Drama Horor Misteri
Sri Asih Alana menemukan kebenaran tentang asalnya: dia bukan manusia biasa. Dia mungkin menjadi hadiah untuk kemanusiaan dan menjadi pelindungnya sebagai Sri Asih. Atau kehancuran, jika dia tidak bisa mengendalikan kemarahannya. Aksi, Petualangan, Fiksi Ilmiah, Fantasi, Drama	Sri Asih Alana menemukan kebenaran tentang asalnya dia bukan manusia biasa Dia mungkin menjadi hadiah untuk kemanusiaan dan menjadi pelindungnya sebagai Sri Asih Atau kehancuran jika dia tidak bisa mengendalikan kemarahannya Aksi Petualangan Fiksi Ilmiah Fantasi Drama

2. Case Folding

Pada proses Case Folding, semua huruf dalam teks atau dokumen diubah menjadi huruf kecil (lowercase). Tabel 3 menampilkan hasil dari Case Folding pada data sampel.

Tabel 3. Hasil Case Folding

Masukkan	Hasil
Pengabdi Setan Communion Setelah pindah dari rumah mereka ke gedung apartemen	pengabdi setan communion setelah pindah dari rumah mereka ke gedung apartemen

Masukkan	Hasil
teror baru menunggu keluarga Rini Drama Horor Misteri	teror baru menunggu keluarga rini drama horor misteri
Sri Asih Alana menemukan kebenaran tentang asalnya dia bukan manusia biasa Dia mungkin menjadi hadiah untuk kemanusiaan dan menjadi pelindungnya sebagai Sri Asih Atau kehancuran jika dia tidak bisa mengendalikan kemarahannya Aksi Petualangan Fiksi Ilmiah Fantasi Drama	sri asih alana menemukan kebenaran tentang asalnya dia bukan manusia biasa dia mungkin menjadi hadiah untuk kemanusiaan dan menjadi pelindungnya sebagai sri asih atau kehancuran jika dia tidak bisa mengendalikan kemarahannya aksi petualangan fiksi ilmiah fantasi drama

3. Tokenization

Tokenization adalah langkah untuk membagi teks menjadi unit-unit yang lebih kecil, dikenal sebagai token. Token ini bisa berupa kata, kalimat, atau karakter. Dalam penelitian ini, token yang digunakan adalah kata. Tabel 4 menampilkan hasil Tokenization dari data sampel.

Tabel 4. Hasil Tokenization

Masukkan	Hasil
pengabdi setan communion setelah pindah dari rumah mereka ke gedung apartemen teror baru menunggu keluarga rini drama horor misteri	["pengabdi", "setan", "communion", "setelah", "pindah", "dari", "rumah", "gedung", "apartemen", "teror", "baru", "menunggu", "keluarga", "rini", "drama", "horor", "misteri"]
sri asih alana menemukan kebenaran tentang asalnya dia bukan manusia biasa dia mungkin menjadi hadiah untuk kemanusiaan dan menjadi pelindungnya sebagai sri asih atau kehancuran jika dia tidak bisa mengendalikan kemarahannya aksi petualangan fiksi ilmiah fantasi drama	["sri", "asih", "alana", "menemukan", "kebenaran", "tentang", "asalnya", "dia", "bukan", "manusia", "biasa", "dia", "mungkin", "menjadi", "hadiah", "untuk", "kemanusiaan", "dan", "menjadi", "pelindungnya", "sebagai", "sri", "asih", "atau", "kehancuran", "jika", "dia", "tidak", "bisa", "mengendalikan", "aksi", "petualangan", "fiksi", "ilmiah", "fantasi", "drama"]

"kemarahannya", "aksi",
"petualangan", "fiksi",
"ilmiah", "fantasi", "drama"]

4. Stopwords Removal

Stopwords Removal merupakan tahap penghapusan kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis. Penghapusan stopwords ini membantu mengurangi noise dan memastikan fokus pada kata-kata yang lebih informatif. Hasil Stopwords Removal dari data sampel dapat dilihat pada Tabel 5.

Tabel 5. Hasil Stopwords Removal

Masukkan	Hasil
["pengabdi", "setan", "communion", "setelah", "pindah", "dari", "rumah", "mereka", "ke", "gedung", "apartemen", "teror", "baru", "menunggu", "keluarga", "rini", "drama", "horor", "misteri"]	["pengabdi", "setan", "communion", "pindah", "rumah", "gedung", "apartemen", "teror", "menunggu", "keluarga", "rini", "drama", "horor", "misteri"]
["sri", "asih", "alana", "menemukan", "kebenaran", "tentang", "asalnya", "dia", "bukan", "manusia", "biasa", "dia", "mungkin", "menjadi", "hadiah", "untuk", "kemanusiaan", "dan", "menjadi", "pelindungnya", "sebagai", "sri", "asih", "atau", "kehancuran", "jika", "dia", "tidak", "bisa", "mengendalikan", "kemarahannya", "aksi", "petualangan", "fiksi", "ilmiah", "fantasi", "drama"]	["sri", "asih", "alana", "menemukan", "kebenaran", "asalnya", "manusia", "hadiah", "kemanusiaan", "pelindungnya", "sri", "asih", "kehancuran", "mengendalikan", "kemarahannya", "aksi", "petualangan", "fiksi", "ilmiah", "fantasi", "drama"]

5. Stemming

Stemming merupakan proses untuk mereduksi kata-kata menjadi bentuk dasar atau akarnya.

Bertujuan untuk menghilangkan variasi bentuk kata dan menyederhanakan analisis. Teknik ini membantu mengurangi kompleksitas dan memastikan konsistensi dalam analisis teks. Hasil Stemming dari data sampel dapat dilihat pada Tabel 6.

Tabel 6. Hasil Stemming

Masukkan	Hasil
["pengabdi", "setan", "communion", "pindah", "rumah", "gedung", "apartemen", "teror", "menunggu", "keluarga", "rini", "drama", "horor", "misteri"]	["pengabdi", "setan", "communion", "pindah", "rumah", "gedung", "apartemen", "teror", "tunggu", "keluarga", "rini", "drama", "horor", "misteri"]
["sri", "asih", "alana", "menemukan", "kebenaran", "asalnya", "manusia", "hadiah", "kemanusiaan", "pelindungnya", "sri", "asih", "kehancuran", "mengendalikan", "kemarahannya", "aksi", "petualangan", "fiksi", "ilmiah", "fantasi", "drama"]	["sri", "asih", "alana", "temu", "benar", "asal", "manusia", "hadiah", "manusia", "lindung", "sri", "asih", "hancur", "mengendalikan", "marah", "aksi", "tualang", "fiksi", "ilmiah", "fantasi", "drama"]

4.3. Perhitungan

Di tahap ini, data yang sudah dibersihkan dan diolah akan dihitung bobot serta similarity-nya guna mendapatkan hasil rekomendasi yang sesuai dengan kata kunci yang diberikan. Kata kunci yang dipakai untuk perhitungan ini adalah "Horor Rumah Setan Hantu".

1. Term Frequency (TF)

Term Frequency (TF) mengukur seberapa sering suatu istilah muncul dalam suatu item (dokumen). Rumus untuk menghitung TF untuk sebuah istilah t dalam sebuah dokumen d adalah:

$$TF(t, d) = \frac{f_{t,d}}{T_d} \quad (1)$$

Keterangan:

$f_{t,d}$: Frekuensi kemunculan istilah t pada dokumen d

T_d : Total jumlah istilah pada dokumen d

Hasil perhitungan Term Frequency dari kata kunci dan sampel film dapat dilihat pada Tabel 7.

Tabel 7. Hasil Perhitungan TF

Kata	Term Frequency		...
	Kata Kunci	Pengabdian Setan 2: Communion	
Horor	0.25	0.07142857142857142	...
Rumah	0.25	0.07142857142857142	...
Setan	0.25	0.07142857142857142	...
Hantu	0.25	0	...
...

2. Inverse Document Frequency (IDF)

Inverse Document Frequency (IDF) mengukur seberapa umum atau langka suatu istilah di antara semua dokumen. Rumus untuk IDF untuk istilah t adalah:

$$IDF(t, D) = \log\left(\frac{N}{n_t + 1}\right) \quad (2)$$

Keterangan:

N : Total jumlah dokumen dalam koleksi

n_t : Jumlah dokumen yang mengandung istilah t

Tabel 8 menunjukkan perhitungan Document Frequency (DF) dan Inverse Document Frequency (IDF). Document Frequency adalah jumlah dokumen dalam korpus di mana istilah tertentu muncul setidaknya satu kali. Penambahan angka 1 pada penyebut mencegah

pembagian dengan nol. Pada percobaan ini, 20 film digunakan sebagai data.

Tabel 8. Perhitungan DF dan IDF

Kata	DF	IDF
Horor	10	0.3010299956639812
Rumah	3	0.8239087409443188
Setan	2	1
Hantu	1	1.3010299956639813
...

3. Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF adalah kombinasi dari TF dan IDF, memberikan bobot yang lebih tinggi pada istilah yang sering muncul dalam dokumen tertentu tetapi jarang di dokumen lainnya. Rumus TF-IDF untuk istilah t dalam dokumen adalah:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t, D) \quad (3)$$

Hasil perhitungan bobot TF-IDF dari kata kunci dan sampel film dapat dilihat di Tabel 9.

Tabel 9. Perhitungan TF-IDF

Kata	TF-IDF		...
	Kata Kunci	Pengabdian Setan 2: Communion	
Horor	0.0752574989159953	0.021502142547427227	...
Rumah	0.2059771852360797	0.058850624353165626	...
Setan	0.25	0.07142857142857142	...
Hantu	0.3252574989159953	0	...
...

4. Cosine Similarity

Cosine Similarity mengukur sudut antara dua vektor. Dalam konteks Content Based Filtering, hitung kesamaan antara dua dokumen yang

direpresentasikan sebagai vektor. Rumus Cosine Similarity antara dua vektor A dan B adalah:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \times \|B\|} \quad (4)$$

Keterangan:

$A \cdot B$: Dot product antara dua vektor

$\|A\|$ dan $\|B\|$: Norm (panjang) dari masing-

masing vektor

Untuk panjang vektor A dapat dihitung dengan rumus sebagai berikut:

$$\|A\| = \sqrt{\sum_{i=1}^n a_i^2} \quad (5)$$

Keterangan:

n : Jumlah komponen dalam vektor A

a_i : Komponen ke- i dari vektor A

Pada Tabel 11, terdapat hasil perhitungan dot product dari dua vektor, yaitu vektor A dan vektor B. Vektor A adalah bobot TF-IDF pada kata kunci, sedangkan Vektor B adalah bobot TF-IDF pada dokumen (film).

Tabel 11. Perhitungan Dot Product

Kata	$A \cdot B$	
	Pengabdi Setan 2: Communion	...
Horor	0.001618197469454581	...
Rumah	0.012121885953650938	...
Setan	0.017857142857142856	...
Hantu	0	...
...
Total	0.03159722628024837	...

Pada Tabel 12, terdapat hasil perhitungan panjang vektor dari kata kunci dan sampel film.

Tabel 12. Perhitungan Panjang Vektor

Kata	Panjang Vektor	
	Kata Kunci	Pengabdi Setan 2: Communion
Horor	0.005663691143091034	0.00046234213412988024
Rumah	0.042426600837778286	0.003463395986757411
Setan	0.0625	0.00510204081632653
Hantu	0.1057924406010887	0
...
Total	0.216382732581958	0.3274815631096327
Akar	0.46516957400711195	0.5722600485003585

Tabel 13 menunjukkan hasil perhitungan Cosine Similarity untuk setiap film.

Tabel 13. Perhitungan Cosine Similarity

No.	Judul Film	Cosine Similarity
1	Sri Asih	0
2	KKN di Desa Penari	0.010277949656955642
3	Miracle in Cell No. 7	0
4	Satria Dewa: Gatotkaca	0
5	Ashiap Man	0
6	The Big 4	0
7	Mendarat Darurat	0
8	Cinta Subuh	0
9	Jailangkung: Sandekala	0.013642053217759516
10	Ivanna	0.009607091124026404
11	Jagat Arwah	0.009370975140788647
12	Ben & Jody	0
13	Inang	0.011809511068126805
14	Nana	0.08648166002115475
15	Pengabdi Setan 2: Communion	0.3017973948727392
16	Menjelang Magrib	0.013764795453227476
17	Like & Share	0
18	Qodrat	0.011045044071225111
19	Qorin	0.012541726359332844
20	PSK Belia	0

4.4. Hasil Rekomendasi

Setelah perhitungan dilakukan, didapatkan tingkat kemiripan (similarity) pada setiap judul film. Tingkat kemiripan ini berguna untuk menentukan

sejauh mana suatu film memiliki kesamaan dengan kata kunci yang diberikan. Semakin tinggi angka kemiripan, semakin dekat kesesuaian antara film dengan kata kunci tersebut. Hasil perhitungan dengan 20 sampel data, yang telah diurutkan dari nilai terkecil ke terbesar, ditampilkan pada Tabel 14.

Tabel 14. Hasil Rekomendasi

No.	Judul Film	Cosine Similarity
1	Pengabdian Setan 2: Communion	0.3017973948727392
2	Nana	0.08648166002115475
3	Menjelang Magrib	0.013764795453227476
4	Jailangkung: Sandekala	0.013642053217759516
5	Qorin	0.012541726359332844
6	Inang	0.011809511068126805
7	Qodrat	0.011045044071225111
8	KKN di Desa Penari	0.010277949656955642
9	Ivanna	0.009607091124026404
10	Jagat Arwah	0.009370975140788647
11	Sri Asih	0
12	Miracle in Cell No. 7	0
13	Satria Dewa: Gatotkaca	0
14	Ashiap Man	0
15	The Big 4	0
16	Mendarat Darurat	0
17	Cinta Subuh	0
18	Ben & Jody	0
19	Like & Share	0
20	PSK Belia	0

Berdasarkan Tabel 14 di atas, judul-judul film yang direkomendasikan adalah “Pengabdian Setan 2: Communion”, “Nana”, “Menjelang Magrib”, “Jailangkung: Sandekala”, “Qorin”, “Inang”, “Qodrat”, “KKN di Desa Penari”, “Ivanna”, dan “Jagat Arwah”. Sistem akan menampilkan maksimal 10 film yang diurutkan berdasarkan tingkat kesamaan tertinggi.

5. Kesimpulan

Berdasarkan percobaan menggunakan 20 sampel film, sistem yang dirancang mampu memberikan rekomendasi film dengan akurasi dan ketepatan yang tinggi sesuai dengan kata kunci yang dimasukkan. Film “Pengabdian Setan 2: Communion” menunjukkan nilai similarity tertinggi sebesar 0.3018, diikuti oleh “Nana”, “Menjelang Magrib”, “Jailangkung: Sandekala”, dan “Qorin” dengan nilai similarity masing-masing 0.0865, 0.0138, 0.0136, dan 0.0125. Meskipun demikian, penelitian ini masih memiliki beberapa kekurangan yang dapat diperbaiki untuk meningkatkan performa sistem rekomendasi ini.

6. Daftar Pustaka

- [1] H. Pratista, *Memahami Film*. Yogyakarta: Homerian Pustaka, 2008.
- [2] “Indonesia.go.id - Tren Positif Film Indonesia,” Indonesia.go.id. Diakses: 2 April 2024.
- [3] W. M. C. Nababan, “Tahun 2023, Penonton Film Indonesia Ditargetkan Pecahkan Rekor Baru - Kompas.id,” Kompas. Diakses: 2 April 2024.
- [4] Y. D. R. Pusparisa, “Industri Film Indonesia Akan Makin Atraktif pada 2024 - Kompas.id,” Kompas. Diakses: 2 April 2024.
- [5] “Streaming is the future of TV, but the abundance of platform choice is overwhelming for viewers | Nielsen,” Nielsen. Diakses: 2 April 2024. [Daring]. Tersedia pada: <https://www.nielsen.com/insights/2022/streaming-is-the-future-of-tv-but-abundance-of-platform-choice-is-overwhelming-for-viewers/>

- [6] Y. Pusparisa, "Platform Menonton Film Indonesia Secara Online," databoks. Diakses: 2 April 2024. [Daring]. Tersedia pada: <https://databoks.katadata.co.id/datapublish/2021/03/30/platform-menonton-film-indonesia-secara-online>
- [7] H. Mutiasari, T. W. Purboyo, dan R. A. Nugrahaeni, "Sistem Rekomendasi Film Menggunakan Metode K-Means Clustering," *e-Proceeding of Engineering*, vol. 8, no. 5, 2021.
- [8] D. Nugraha, T. W. Purboyo, dan R. A. Nugrahaeni, "Sistem Rekomendasi Film Menggunakan Metode User Based Collaborative Filtering," *e-Proceeding of Engineering*, vol. 8, no. 5, 2021.
- [9] A. A. Huda, R. Fajarudin, dan A. Hadinegoro, "Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Des 2022, doi: 10.47065/bits.v4i3.2511.
- [10] Joni, Andy, dan K. Wibowo, "Perancangan Website Rekomendasi Film Dengan Menggunakan Metode User Based Collaborative Filtering," *Jurnal Ilmiah Teknik Informatika*, vol. 1, no. 2, hlm. 37–43, Okt 2021, [Daring]. Tersedia pada: <http://ojs.fikom-methodist.net/index.php/>
- [11] E. Ryana Agustian, Munir, dan E. Prasetyo Nugroho, "Sistem Rekomendasi Film Menggunakan Metode Collaborative Filtering dan K-Nearest Neighbors," *Jurnal Aplikasi dan Teori Ilmu Komputer*, vol. 3, no. 1, Mar 2020, Diakses: 10 Mei 2024. [Daring]. Tersedia pada: <https://ejournal.upi.edu/index.php/JATIKOM>
- [12] V. Atina dan D. Hartanti, "KNOWLEDGE BASED RECOMMENDATION MODELING FOR CLOTHING PRODUCT SELECTION RECOMMENDATION SYSTEM," *Jurnal Teknik Informatika (Jutif)*, vol. 3, no. 5, hlm. 1407–1413, Okt 2022, doi: 10.20884/1.jutif.2022.3.5.584.
- [13] D. Roy dan M. Dutta, "A systematic review and research perspective on recommender systems," *J Big Data*, vol. 9, no. 1, Des 2022, doi: 10.1186/s40537-022-00592-5.
- [14] D. A. Putri, D. Pramesti, D. I, dan W. Santiyasa, "Penerapan Metode Content-Based Filtering dalam Sistem Rekomendasi Video Game," *JNATIA*, vol. 1, no. 1, 2022.
- [15] R. Huang, "Improved content recommendation algorithm integrating semantic information," *J Big Data*, vol. 10, no. 1, Des 2023, doi: 10.1186/s40537-023-00776-7.
- [16] R. W. Pratiwi dan Y. S. Nugroho, "Prediksi Rating Film Menggunakan Metode Naïve Bayes," *Duta.com: Jurnal Ilmiah Teknologi Informasi dan Komunikasi*, vol. 12, no. 1, hlm. 91–108, 2017, [Daring]. Tersedia pada: <https://www.kaggle.com>
- [17] S. Rosetya Wardhana dan R. Kembang Hapsari, "Sistem Rekomendasi Film dengan Menggunakan Pendekatan Collaborative Filtering Berdasarkan Class," *Prosiding Seminar Implementasi Teknologi Informasi dan Komunikasi*, vol. 2, no. 1, 2023, doi: 10.31284/p.semtik.2023-1.4153.