

**IMPLEMENTASI METODE *CONTENT BASED FILTERING* PADA
SISTEM REKOMENDASI FILM**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

STEFANUS DWI CHRISTIANTO

22.11.4982

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

**IMPLEMENTASI METODE *CONTENT BASED FILTERING*
PADA SISTEM REKOMENDASI FILM**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

STEFANUS DWI CHRISTIANTO

22.11.4982

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

HALAMAN PERSETUJUAN

SKRIPSI

**IMPLEMENTASI METODE *CONTENT BASED FILTERING*
PADA SISTEM REKOMENDASI FILM**

yang disusun dan diajukan oleh

Stefanus Dwi Christianto

22.11.4982

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 15 Desember 2025

Dosen Pembimbing,



**Dr. Emigawaty, M.Kom
NIK. 19030xxxx**

HALAMAN PENGESAHAN
SKRIPSI
IMPLEMENTASI METODE *CONTENT BASED FILTERING*
PADA SISTEM REKOMENDASI FILM

yang disusun dan diajukan oleh

Stefanus Dwi Christianto

22.11.4982

Telah dipertahankan di depan Dewan Penguji
pada tanggal 15 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ali Mustopa, S.Kom., M.Kom.
NIK. 190302192



Nur Aini, A.Md., S.Kom., M.Kom.
NIK. 190302066



Dr. Emigawaty, M.Kom
NIK. 190302226



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 15 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Stefanus Dwi Christianto
NIM : 22.11.4982

Menyatakan bahwa Skripsi dengan judul berikut:

Implementasi Metode *Content-Based Filtering* pada Sistem Rekomendasi Film

Dosen Pembimbing : Dr. Emigawaty, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 15 Desember 2025

Yang Meny:


Stefanus Dwi Christianto

HALAMAN PERSEMBAHAN

Karya tulis ini dipersembahkan sebagai wujud rasa hormat dan terima kasih kepada :

1. Kedua Orant Tua, Bapak Sutrisno dan Ibu Eli, atas segala dukungan moril, material, serta kasih sayang yang tidak terhingga selama proses pendidikan ini.
2. Kakak saya, Alberth atas segala dukungan dan bantuan baik itu moril atau material selama proses pendidikan ini.
3. Dosen Pembimbing, Ibu Dr. Emigawaty, M.Kom, atas ilmu yang diberikan dalam membimbing penyusunan skripsi ini hingga tuntas.
4. Diri sendiri, atas kerja keras, ketekunan, dan kemampuan untuk tetap bertahan menyelesaikan tanggung jawab akademis ini.
5. Almamater, Universitas Amikom Yogyakarta

KATA PENGANTAR

Puji syukur penulis panjatkan ke Tuhan Yang Maha Esa, karena atas berkat, rahmat, dan karunia-Nya, penulis dapat menyelesaikan skripsi dengan judul **“Implementasi Model *Content-Based Filtering* pada Sistem Rekomendasi Film”**. Penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Program Studi Informatika, Universitas Amikom Yogyakarta.

Penulis menyadari bahwa kelancaran penyusunan skripsi ini tidak lepas dari doa, dukungan, dan bimbingan orang-orang terdekat. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih yang setulus-tulusnya kepada:

1. Kedua Orang Tuaku, Bapak Sutrisno dan Ibu Eli, yang senantiasa memberikan kasih sayang, doa yang tiada henti, serta dukungan moril maupun materil yang luar biasa. Kalian adalah motivasi terbesar penulis untuk menyelesaikan pendidikan ini.
2. Kakak Alberth, terima kasih atas semangat, bantuan, dan nasihat yang selalu diberikan selama masa perkuliahan hingga penyelesaian skripsi ini.
3. Ibu Dr. Emigawaty, M.Kom, selaku Dosen Pembimbing, terima kasih banyak atas kesabaran, waktu, dan ilmu yang telah diberikan dalam membimbing, mengarahkan, serta memotivasi penulis selama proses penyusunan skripsi ini hingga tuntas.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan.

Yogyakarta, 15 Desember 2025

Penulis

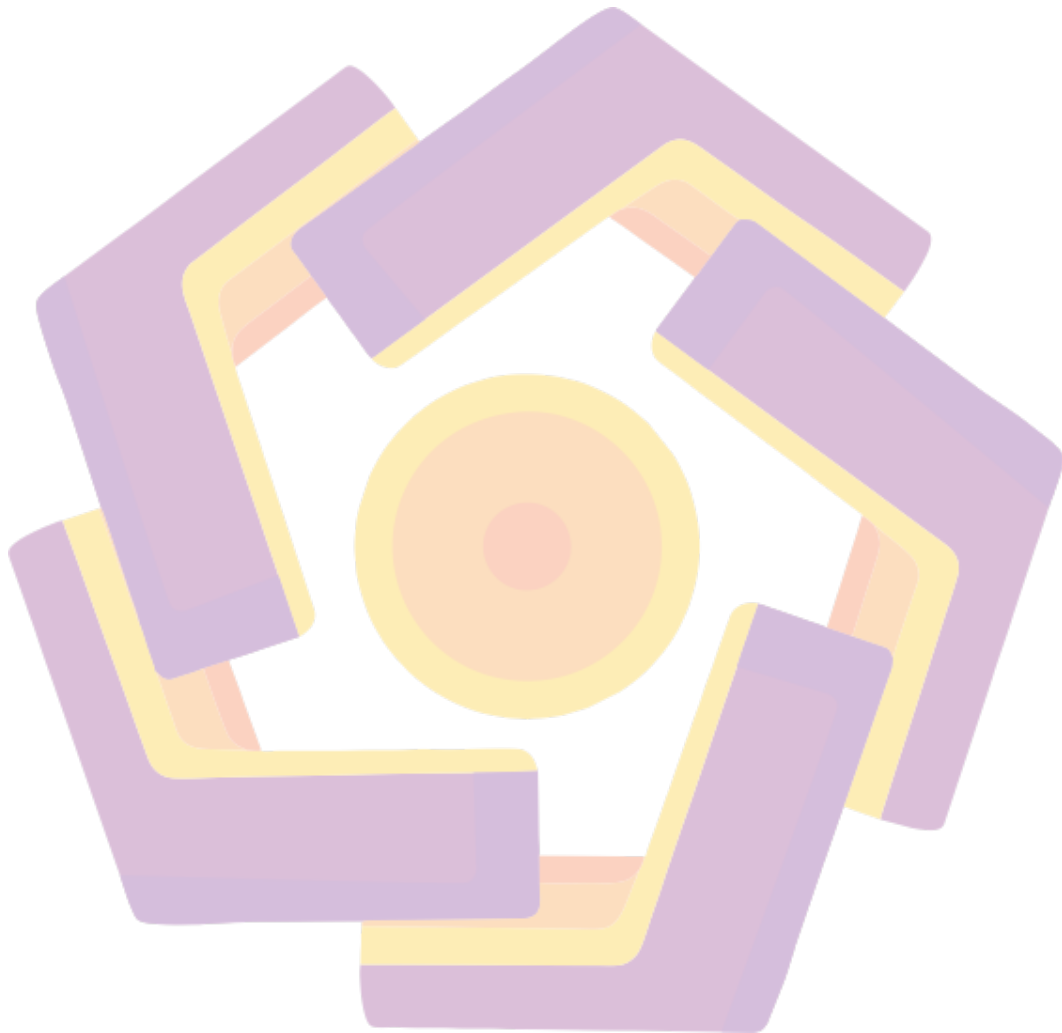
DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
<i>ABSTRACT</i>	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori.....	15
2.2.1 Sistem Rekomendasi.....	15
2.2.2 Content Based Filtering	15
2.2.3 Vector Space Model (VSM)	15
2.2.4 Text Preprocessing.....	16
2.2.5 Term Frequency – Inverse Document Frequency (TF-IDF).....	16
2.2.6 Cosine Similarity.....	16
2.2.7 Black Box Testing.....	17
BAB III METODE PENELITIAN	18

3.1 Objek Penelitian	18
3.2 Alur Penelitian	18
3.3 Alat dan Bahan	21
3.3.1 Data Penelitian	21
3.3.2 Alat/Instrumen Penelitian	22
BAB IV HASIL DAN PEMBAHASAN	25
4.1 Implementasi Sistem	25
4.1.1 Pra-Pemrosesan Data	29
4.1.2 Analisis Data Eksploratif (EDA)	25
4.1.3 Rekayasa Fitur (Fitur Engineering).....	33
4.1.4 Implementasi Model Sistem Rekomendasi.....	34
4.1.5 Pengujian dan Hasil Rekomendasi.....	35
4.1.6 Evaluasi Kinerja Model Rekomendasi Film	36
4.2 Implementasi Prototipe Aplikasi Website	38
4.2.1 Lingkungan Implementasi	38
4.2.2 Antarmuka Pengguna (User Interface)	39
4.2.3 Pengujian Fungsionalitas Website (<i>Blackbox Testing</i>).....	40
4.2.4 Deployment Prototipe Aplikasi.....	42
4.3 Pembahasan Hasil	43
4.3.1 Analisis Akurasi Dan Relevansi Rekomendasi/.....	43
4.3.2 Evaluasi Kualitas (Validasi Metrik).....	44
4.3.3 Analisis Keterbatasan.....	44
BAB V PENUTUP	46
5.1 Kesimpulan	46
5.2 Saran.....	47
REFERENSI	49

DAFTAR TABEL

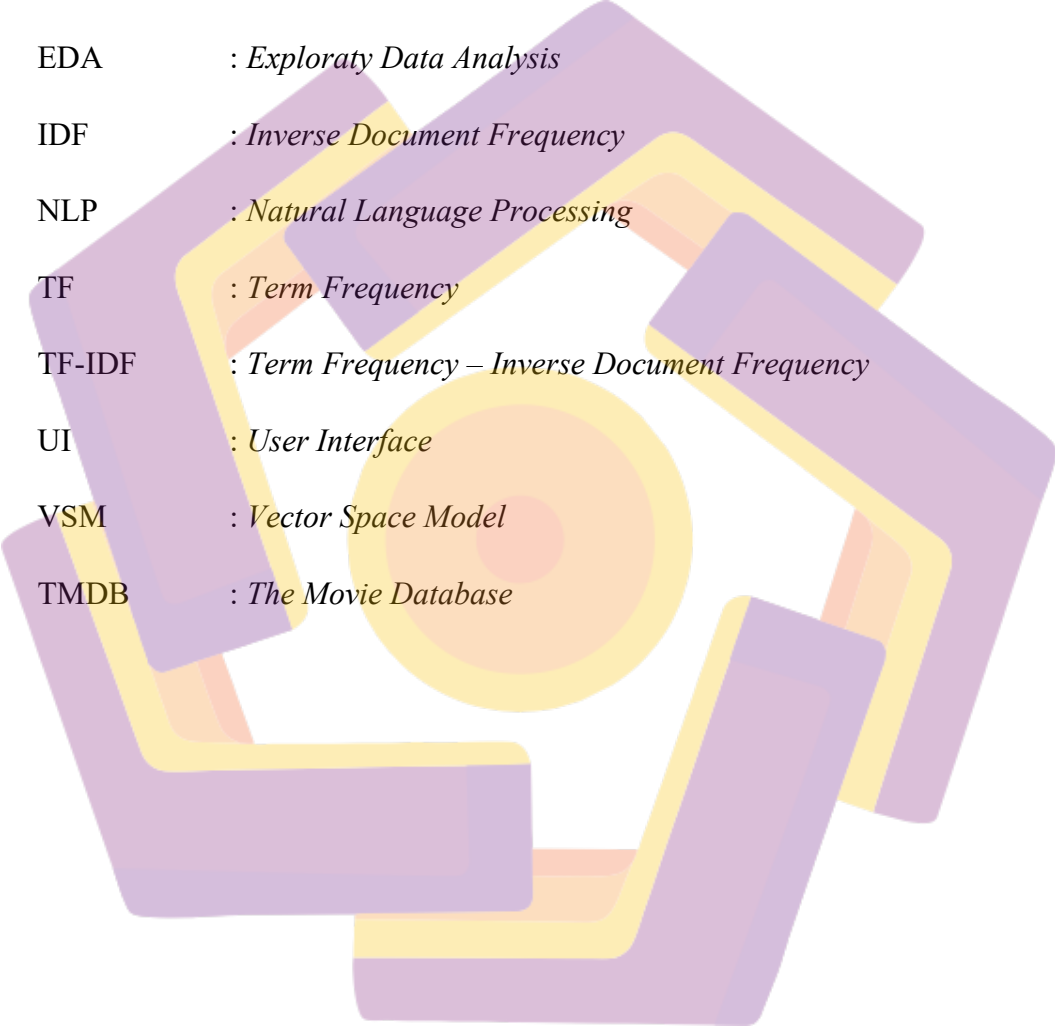
Tabel 2. 1 Keaslian Penelitian	11
Tabel 4.1 Uji Hasil Rekomendasi	37
Tabel 4.2 Blackbox Testing	40



DAFTAR GAMBAR

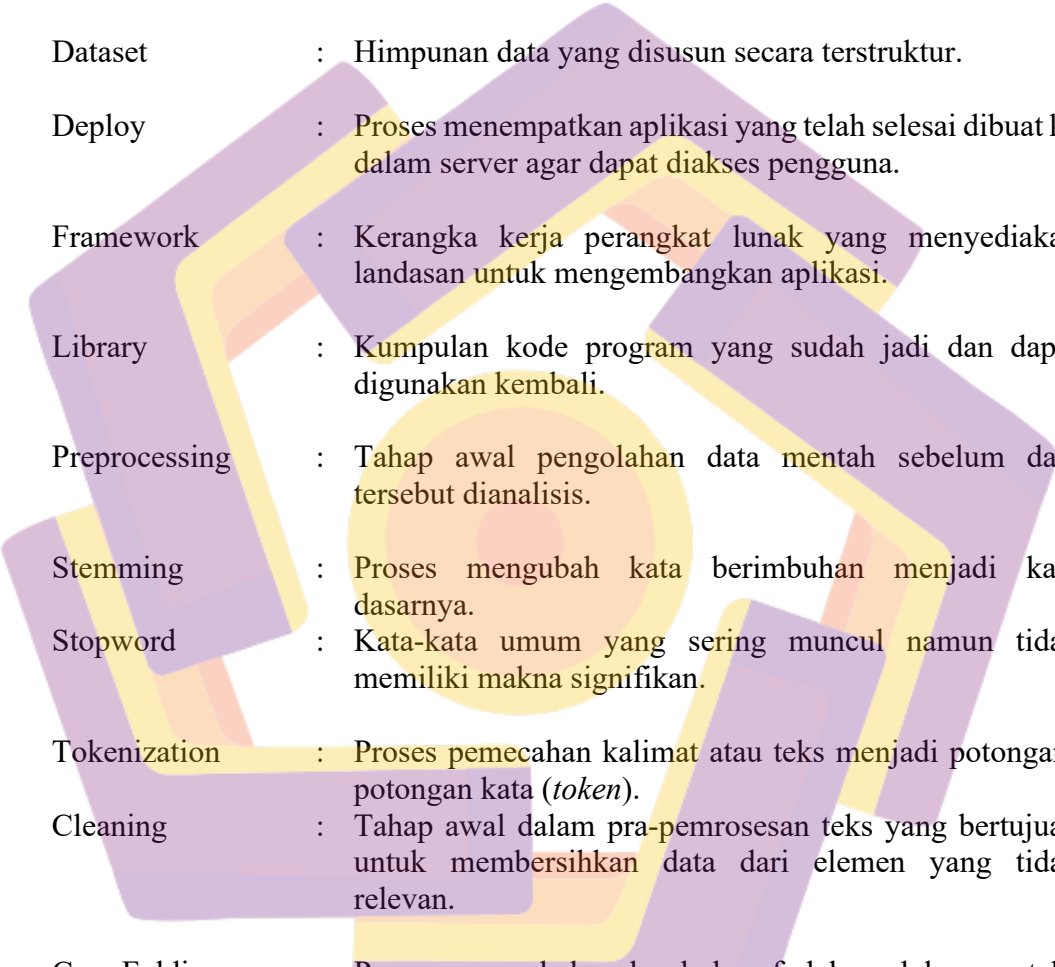
Gambar 3.1 Alur Penelitian	19
Gambar 4.1 Analisis Tahun Rilis	25
Gambar 4.2 Analisis Genre	26
Gambar 4.3 Analisis Genre Sutradara dan Aktor	27
Gambar 4. 4 Analisis Popularity	28
Gambar 4.5 Informasi Awal Dataset	29
Gambar 4.6 Mengatasi Missing Values	29
Gambar 4.7 Penyaringan Data	30
Gambar 4. 8 Genre Setelah Filtering	31
Gambar 4. 9 Sutradara dan Pemeran Setelah Filtering	32
Gambar 4.10 Pembentukan Fitur Gabungan	33
Gambar 4.11 Perhitungan TF-IDF & Cosine Similarity	34
Gambar 4.12 Fungsi Rekomendasi Film	34
Gambar 4.13 Pengujian Rekomendasi Film	35
Gambar 4.14 Fungsi Uji Hasil Rekomendasi	36
Gambar 4.16 Halaman Rekomendasi Film	39
Gambar 4.17 Hasil Rekomendasi Film	39
Gambar 4.18 Analisis Hasil Rekomendasi dan Evaluasi Metrik Film “Harry Potter”	43

DAFTAR LAMBANG DAN SINGKATAN



CBF	: <i>Content-Based Filtering</i>
CF	: <i>Collaborative Filtering</i>
CSV	: <i>Comma-Separated Values</i>
EDA	: <i>Exploraty Data Analysis</i>
IDF	: <i>Inverse Document Frequency</i>
NLP	: <i>Natural Language Processing</i>
TF	: <i>Term Frequency</i>
TF-IDF	: <i>Term Frequency – Inverse Document Frequency</i>
UI	: <i>User Interface</i>
VSM	: <i>Vector Space Model</i>
TMDB	: <i>The Movie Database</i>

DAFTAR ISTILAH



Cold Start	: Situasi di mana sistem rekomendasi kesulitan memberikan saran karena kurangnya data interaksi.
Cosine Similarity	: Metode untuk mengukur tingkat kemiripan antara dua vektor dalam ruang multidimensi.
Dataset	: Himpunan data yang disusun secara terstruktur.
Deploy	: Proses menempatkan aplikasi yang telah selesai dibuat ke dalam server agar dapat diakses pengguna.
Framework	: Kerangka kerja perangkat lunak yang menyediakan landasan untuk mengembangkan aplikasi.
Library	: Kumpulan kode program yang sudah jadi dan dapat digunakan kembali.
Preprocessing	: Tahap awal pengolahan data mentah sebelum data tersebut dianalisis.
Stemming	: Proses mengubah kata berimbuhan menjadi kata dasarnya.
Stopword	: Kata-kata umum yang sering muncul namun tidak memiliki makna signifikan.
Tokenization	: Proses pemecahan kalimat atau teks menjadi potongan-potongan kata (<i>token</i>).
Cleaning	: Tahap awal dalam pra-pemrosesan teks yang bertujuan untuk membersihkan data dari elemen yang tidak relevan.
Case Folding	: Proses mengubah seluruh huruf dalam dokumen teks menjadi bentuk huruf kecil (<i>lowercase</i>) untuk menyeragamkan data dan menghindari duplikasi kata akibat perbedaan kapitalisasi.

INTISARI

Perkembangan pesat platform *streaming* film di era digital telah menyebabkan lonjakan volume konten yang masif, sehingga memicu fenomena *information overload*. Kondisi ini berdampak negatif pada pengalaman pengguna, mengakibatkan kelelahan dalam pengambilan keputusan (*decision fatigue*) dan inefisiensi waktu dalam menemukan tontonan yang relevan dengan preferensi mereka. Penelitian ini bertujuan mengatasi permasalahan tersebut dengan membangun sebuah sistem rekomendasi film menggunakan metode *Content-Based Filtering*. Metodologi penyelesaian masalah dilakukan melalui beberapa tahapan sistematis, meliputi akuisisi dataset film, pra-pemrosesan data teks, rekayasa fitur gabungan (*soup*) yang mencakup genre dan sinopsis, pembobotan fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), serta perhitungan tingkat kemiripan antar dokumen menggunakan algoritma *Cosine Similarity*. Sistem yang dirancang kemudian diimplementasikan ke dalam sebuah prototipe aplikasi berbasis web fungsional. Hasil akhir penelitian menunjukkan bahwa sistem mampu memberikan rekomendasi yang akurat dan relevan, dibuktikan dengan evaluasi kinerja yang mencapai persentase kesamaan genre hingga 80% dan kesamaan kru film sebesar 70% pada skenario pengujian. Penelitian ini memberikan manfaat praktis bagi pengguna dalam mempercepat pencarian film, serta menjadi rujukan bagi penyedia layanan hiburan untuk meningkatkan keterlibatan pengguna melalui fitur personalisasi. Pengembangan selanjutnya disarankan untuk mengintegrasikan metode *Hybrid Filtering* guna memperkaya variasi rekomendasi.

Kata kunci: Sistem Rekomendasi, *Content-Based Filtering*, TF-IDF, *Cosine Similarity*, Film.

ABSTRACT

The rapid development of movie streaming platforms in the digital era has caused a massive surge in content volume, triggering the phenomenon of information overload. This condition negatively impacts user experience, leading to decision fatigue and time inefficiency in finding content relevant to their preferences. This study aims to address this issue by developing a movie recommendation system using the Content-Based Filtering method. The problem-solving methodology involves several systematic stages, including dataset acquisition, text data preprocessing, combined feature engineering (soup) encompassing genres and synopses, feature weighting using Term Frequency-Inverse Document Frequency (TF-IDF), and similarity calculation using the Cosine Similarity algorithm. The designed system was subsequently implemented into a functional web-based application prototype. The results indicate that the system is capable of providing accurate and relevant recommendations, evidenced by performance evaluations achieving a genre similarity percentage of up to 80% and film crew similarity of 70% in testing scenarios. This research offers practical benefits for users by accelerating the movie search process and serves as a reference for entertainment providers to enhance user engagement through personalization features. Future development is recommended to integrate Hybrid Filtering methods to enrich recommendation variety.

Keyword: Recommendation System, Content-Based Filtering, TF-IDF, Cosine Similarity, Movie.