



**PENERAPAN ALGORITMA *CONTENT BASED FILTERING* UNTUK
SISTEM REKOMENDASI FILM DAN LAYANAN *STREAMING FILM***

SKRIPSI

ADE KURNIA PUTRA

NIM. 2110511053

S1 INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

2025



**PENERAPAN ALGORITMA *CONTENT BASED FILTERING* UNTUK
SISTEM REKOMENDASI FILM DAN LAYANAN *STREAMING FILM***

SKRIPSI

**Diajukan Sebagai salah satu syarat untuk Memperoleh Gelar Sarjana
Komputer**

ADE KURNIA PUTRA

NIM. 2110511053

S1 INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

2025

PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya sendiri dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Ade Kurnia Putra

NIM : 2110511053

Tanggal : 7 Juli 2025

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 7 Juli 2025

Yang Menyatakan



Ade Kurnia Putra

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademika Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Ade Kurnia Putra

NIM : 2110511053

Fakultas : Ilmu Komputer

Program Studi : S-1 Informatika

Demi Pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (Non – exclusive Royalty Free Right) atas skripsi saya yang berjudul:

**PENERAPAN ALGORITMA *CONTENT BASED FILTERING* UNTUK
SISTEM REKOMENDASI FILM DAN LAYANAN *STREAMING FILM***

Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/memformatkan, mengelola dalam bentuk pangkalan data (basis data), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 7 Juli 2025

Yang menyatakan,



Ade Kurnia Putra

LEMBAR PENGESAHAN

Judul : Penerapan Algoritma *Content Based Filtering* Untuk Sistem
Rekomendasi Film Dan Layanan *Streaming* Film
Nama : Ade Kurnia Putra
NIM : 2110511053
Program Studi : SI Informatika

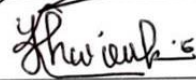
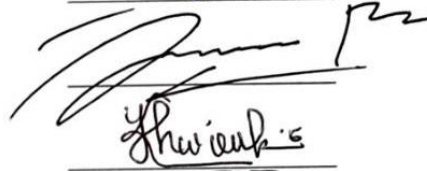
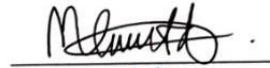
Disetujui oleh:

Penguji 1:
Nurhafifah Matondang, S.Kom., M.M., M.T.I.

Penguji 2:
Nurul Afifah Arifuddin, S.Pd., M.T.

Pembimbing 1:
Indra Permana Solihin, S.Kom., M.Kom.

Pembimbing 2:
Kharisma Wiati Gusti, M.T.



Diketahui oleh:

Koordinator Program Studi:
Dr. Widya Cholil, M.I.T.
NIP. 2211122080
Dekan Fakultas Ilmu Komputer:
Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM
NIP. 197605082003121002



Tanggal Ujian Tugas Akhir:
4 Juli 2025

KATA PENGANTAR

Alhamdulillah, puji dan syukur peneliti panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat, kesehatan, dan segala karunia-Nya sehingga proses penyusunan dan penyelesaian skripsi tugas akhir dengan judul “Penerapan Algoritma *Content based filtering* Untuk Sistem Rekomendasi Film Dan Layanan *Streaming Film*” dapat berjalan dengan baik.

Skripsi tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Fakultas Ilmu Komputer, Program Studi Informatika, Universitas Pembangunan Nasional “Veteran” Jakarta. Dalam proses penyusunannya, peneliti mendapatkan berbagai bentuk dukungan, bimbingan, bantuan, dan kerja sama dari berbagai pihak, baik dalam aspek moral, material, maupun spiritual, yang sangat membantu dalam penyelesaian skripsi tugas akhir ini.

Oleh karena itu, peneliti ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Allah SWT.
2. Bapak, Ibu, dan Kakak peneliti yang senantiasa memberikan dukungan, dan semangat dalam menyelesaikan skripsi tugas akhir.
3. Bapak Prof. Dr. Ir. Supriyanto, M.Sc., IPM. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
4. Ibu Dr. Widya Cholil, M.I.T. selaku Ketua Program Studi S1 Informatika.
5. Bapak Musthofa Galih Pradana, M.Kom selaku dosen pembimbing akademik.
6. Bapak Indra Permana Solihin, S.Kom.,M.Kom sebagai pembimbing pertama proposal dan skripsi, atas bimbingan, motivasi, serta saran konstruktif yang diberikan.
7. Ibu Kharisma Wiati Gusti, M.T. sebagai pembimbing kedua proposal dan skripsi, atas bimbingan, motivasi, serta saran konstruktif yang diberikan.
8. Bapak/Ibu Dosen Fakultas Ilmu Komputer yang telah dengan penuh dedikasi mengajar dan berbagi pengetahuan, sehingga peneliti dapat memperdalam wawasan dan ilmu selama menjadi mahasiswa di Fakultas Ilmu Komputer.

9. Seluruh teman-teman yang sudah membantu, mendukung, mengajak, memberikan motivasi, serta mendoakan selama proses penelitian dan penyusunan skripsi tugas akhir.
10. Semua pihak yang tidak dapat disebutkan satu per satu atas semua kontribusi, baik secara langsung maupun tidak langsung, dalam membantu peneliti menyelesaikan skripsi tugas akhir ini.

Penulis menyadari bahwa skripsi tugas akhir ini masih memiliki kekurangan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun untuk perbaikan penelitian ini di masa depan. Semoga skripsi tugas akhir ini dapat memberikan kontribusi yang berharga bagi pembaca, baik dalam memperluas wawasan maupun menambah pemahaman di bidang yang dibahas. Selain itu, semoga hasil dari skripsi tugas akhir ini dapat bermanfaat dan menjadi referensi yang bernilai bagi berbagai kalangan.

Tangerang Selatan, 22 Juni 2025



Ade Kurnia Putra

PENERAPAN ALGORITMA CONTENT BASED FILTERING UNTUK SISTEM REKOMENDASI FILM DAN LAYANAN *STREAMING* FILM

Ade Kurnia Putra

ABSTRAK

Banyaknya pilihan film dan layanan *streaming* seringkali menyulitkan pengguna dalam pengambilan keputusan. Penelitian ini bertujuan untuk mengatasi kesulitan tersebut dengan merancang sebuah sistem rekomendasi film dan layanan *streaming* film berbasis *content-based filtering*. Data film diperoleh dengan teknik *scraping* dari *API The Movie Database* dengan jumlah 15.813 data film. Kemudian data tersebut melalui tahapan *preprocessing* dengan tahapan *case folding*, *cleaning*, *tokenizing*, dan *stemming*. Selanjutnya model dibuat menggunakan teknik *CountVectorizer* untuk mengubah fitur konten tekstual film menjadi vektor dan *Cosine Similarity* untuk mengukur kemiripannya. Model yang sudah dibuat akan di evaluasi menggunakan metrik evaluasi *MAP@K* yang mendapatkan skor terbesar pada *MAP@1* dengan skor 0.7577. Kemudian model diintegrasikan pada *website* yang dibuat menggunakan framework *ReactJs* sebagai *frontend* dan *Flask* sebagai *backend*. Fungsionalitas *website* diuji dengan *black box testing* dengan hasil berhasil pada semua skenario. Selanjutnya penilaian pengguna diuji menggunakan *user acceptance testing* dengan metode *beta testing* yang mendapatkan skor akhir 4,58 dari skala 5. Hasil ini membuktikan bahwa pendekatan *content-based filtering* dengan *CountVectorizer* dan *Cosine Similarity* terbukti efektif dan cocok untuk membangun sistem rekomendasi film dan layanan *streaming* film yang personal dan memuaskan bagi pengguna.

Kata Kunci: *Content based filtering*, *CountVectorizer*, *Cosine Similarity*, Film, Layanan *Streaming* Film

IMPLEMENTATION OF CONTENT BASED FILTERING ALGORITHM FOR MOVIE RECOMMENDATION SYSTEM AND MOVIE STREAMING SERVICES

Ade Kurnia Putra

ABSTRACT

The abundance of movies and streaming services often hinders user decision-making. This research aims to overcome these difficulties by designing a movie recommendation system and movie streaming services based on content-based filtering. Movie data is obtained by scraping from The Movie Database API with a total of 15,813 data. The data goes through preprocessing stages of case folding, cleaning, tokenizing, and stemming. Model is created using CountVectorizer technique to convert the film's textual content into vectors and Cosine Similarity to measure the similarity. The model is evaluated using MAP@K metrics which gets the largest score on MAP@1 with a score of 0,7577. The model is integrated on the website created using ReactJs framework as the frontend and Flask as the backend. The website functionality was tested with black box testing with successful results in all scenarios. Furthermore, user assessment was tested using user acceptance testing (UAT) with beta testing method which received a final score of 4.58 out of 5. These results prove that content-based filtering approach with CountVectorizer and Cosine Similarity is effective and suitable for building movie recommendation systems and movie streaming services that are personalized and satisfying for users.

Keywords: Content based filtering, CountVectorizer, Cosine Similarity, Movie, Movie Streaming Services

DAFTAR ISI

PERNYATAAN ORISINALITAS	ii
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	iii
LEMBAR PENGESAHAN	iv
KATA PENGANTAR	v
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN.....	xv
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	5
1.6 Luaran Penelitian	5
1.7 Sistematika Penulisan	5
BAB 2. TINJAUAN PUSTAKA	7
2.1 Website	7
2.2 Sistem Rekomendasi.....	7
2.3 Film	8
2.4 Layanan <i>Streaming</i> Film.....	8
2.5 <i>Content-based filtering</i>	9
2.5.1 CountVectorizer.....	10
2.5.2 Consine Similarity	10

2.6	Preprocessing Data.....	11
2.7	Metrik Evaluasi	12
2.7.1	<i>Precision@k</i>	13
2.7.2	<i>Mean Average Precision at K (MAP@K)</i>	13
2.8	<i>Machine learning</i>	14
2.9	Extreme Programming	15
2.10	The Movie Database (TMDB)	17
2.11	ReactJS.....	18
2.12	Flask.....	18
2.13	Python	18
2.14	Visual Studio Code	19
2.15	Black box testing.....	19
2.16	User Acceptance Tesing.....	20
2.17	Penelitian Terdahulu	20
BAB 3. METODE PENELITIAN		24
3.1	Tahapan Penelitian.....	24
3.1.1	Identifikasi Masalah.....	24
3.1.2	Studi Literatur	25
3.1.3	Pengumpulan Data	25
3.1.4	Preprocesssing Data	26
3.1.5	Pembuatan Model	27
3.1.6	Evaluasi Model	30
3.1.7	Pembuatan <i>Website</i>	32
3.1.8	Integrasi <i>Website</i>	33
3.1.9	Pengujian.....	33
3.1.10	Release	34
3.1.11	Dokumentasi.....	35
3.2	Metode Analisis	35
3.3	Metode Pengembangan	37
3.4	Alat Pendukung Penelitian.....	38
3.5	Jadwal Penelitian	38

BAB 4. HASIL DAN PEMBAHASAN	40
4.1 Hasil Pengumpulan Data.....	40
4.1.1 <i>Scraping</i> Data.....	41
4.1.2 Struktur Penyimpanan di Database	44
4.2 Hasil Preprocessing Data	45
4.2.1 Hasil <i>Case Folding</i>	46
4.2.2 Hasil <i>Cleaning</i>	46
4.2.3 Hasil <i>Tokenizing</i>	49
4.2.4 Hasil <i>Stemming</i>	49
4.3 Hasil Pembuatan Model	50
4.4 Hasil Evaluasi Model	54
4.4.1 Skenario Pertama	55
4.4.2 Skenario Kedua	56
4.5 Hasil Pembuatan <i>Website</i>	58
4.5.1 Home Page	60
4.5.2 Detail Page	62
4.5.3 Backend.....	63
4.6 Hasil Integrasi <i>Website</i>	65
4.7 Hasil Pengujian	65
4.7.1 Hasil <i>Black box testing</i>	66
4.7.2 Hasil <i>User acceptance testing</i>	69
4.8 Release	74
BAB 5. PENUTUP	75
5.1 Kesimpulan	75
5.2 Saran	76
DAFTAR PUSTAKA	78
LAMPIRAN.....	83

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi algoritma <i>Content based filtering</i>	9
Gambar 2.2 Ilustrasi teknik <i>Cosine Similarity</i> (Hung dan Pramanik 2023)	11
Gambar 2.3 Alur kerja metode <i>extreme programming</i> (Sari <i>et al.</i> 2022).....	15
Gambar 3.1 Tahapan penelitian	24
Gambar 3.2 Alur pembuatan model <i>content-based filtering</i>	28
Gambar 3.3 ilustrasi <i>ground truth</i>	36
Gambar 3.4 Ilustrasi hasil rekomendasi model.....	37
Gambar 4.1 Membuat API Key TMDB.....	41
Gambar 4.2 Alur program <i>scraping</i> data	42
Gambar 4.3 Tampilan program <i>scraping</i> saat berjalan.....	44
Gambar 4.4 Contoh data film.....	45
Gambar 4.5 Hasil rekomendasi film	53
Gambar 4.6 Trend penurunan nilai <i>MAP@K</i>	57
Gambar 4.7 Use case <i>website</i>	58
Gambar 4.8 Activity diagram <i>website</i>	59
Gambar 4.9 Halaman <i>home page</i>	60
Gambar 4.10 Delay tombol “Buat Rekomendasi”	60
Gambar 4.11 Tampilan rekomendasi	61
Gambar 4.12 Tombol <i>dislike</i>	62
Gambar 4.13 Halaman <i>detail page</i>	62
Gambar 4.14 Logika rekomendasi layanan <i>streaming</i> film.....	64
Gambar 4.15 <i>Release website</i> di platform <i>GitHub</i>	74

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu	20
Tabel 3.1 Tabel hasil <i>preprocessing</i>	27
Tabel 3.2 Contoh matrix fitur <i>CountVectorizer</i>	29
Tabel 4.1 Attribut database	44
Tabel 4.2 Contoh hasil <i>case folding</i>	46
Tabel 4.3 Hasil <i>handle missing values</i>	47
Tabel 4.4 Contoh hasil pembersihan karakter	47
Tabel 4.5 Contoh hasil <i>stopwords removal</i>	48
Tabel 4.6 Contoh hasil <i>tokenizing</i>	49
Tabel 4.7 Contoh hasil <i>stemming</i>	50
Tabel 4.8 Pembuatan <i>tags</i> pada skenario 1	51
Tabel 4.9 Pembuatan <i>tags</i> pada skenario 2	51
Tabel 4.10 Hasil evaluasi dengan kolom <i>genres</i>	55
Tabel 4.11 Hasil evaluasi tanpa kolom <i>genres</i>	56
Tabel 4.12 Hasil pengujian <i>black box testing</i>	66
Tabel 4.13 Skala <i>likert</i>	70
Tabel 4.14 Hasil kuesioner UAT	70

DAFTAR RUMUS

2.1. <i>Cosine Similarity</i>	11
2.2. <i>Precision@k</i>	13
2.3. <i>Mean Average Precision@k (MAP@k)</i>	14
2.4. <i>Rata-Rata User acceptance testing (UAT)</i>	20
2.5. <i>Nilai Total UAT</i>	20
3.1. <i>Contoh Cosine Similarity</i>	30

DAFTAR LAMPIRAN

Lampiran 1. Survei Kuesioner Penelitian	83
Lampiran 2. Hasil Kuesioner <i>User acceptance testing</i> (UAT).....	87
Lampiran 3. Hasil <i>Black box testing</i>	93
Lampiran 4. Hasil Turnitin	94