

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan era digital telah mengakibatkan lonjakan konten yang masif pada platform *streaming* film di Indonesia, sehingga menimbulkan fenomena *information overload* [1][2]. Akibatnya, pengguna menghadapi kesulitan dalam memilih tontonan yang relevan, sehingga berdampak negatif pada pengalaman mereka. Situasi ini memicu kelelahan dalam pengambilan keputusan (*decision fatigue*) dan membuat proses seleksi menjadi tidak efisien. Fenomena ini juga selaras dengan laporan Nielsen tahun 2022 yang berjudul "*Nielsen's State of Play report reveals that streaming is the future, but consumers are currently overwhelmed by choice*" yang mengonfirmasi bahwa konsumen saat ini merasa terbebani oleh melimpahnya pilihan yang tersedia. Menyebabkan pengguna dapat merasa frustrasi, dan penyedia layanan berisiko kehilangan pelanggan, sehingga saat ini dibutuhkan solusi teknologi yang mampu mempersonalisasi pilihan secara efektif.

Sistem rekomendasi dapat berguna sebagai solusi untuk menyaring banyaknya konten, dan salah satu metode fundamental yang terbukti berhasil adalah *Content-Based Filtering* (CBF). Pendekatan ini bekerja dengan cara merekomendasikan film berdasarkan kemiripan atribut atau konten (seperti genre, sutradara, atau sinopsis) dengan film yang pernah disukai pengguna sebelumnya [3]. Keunggulan utama *Content-Based Filtering* (CBF) adalah kemampuannya memberikan rekomendasi yang spesifik tanpa memerlukan data dari pengguna lain, sehingga sangat relevan untuk personalisasi.

Keefektifan metode CBF telah divalidasi dalam berbagai penelitian di Indonesia. Studi oleh Velamentosa dan Zuliarso (2025) menunjukkan bahwa implementasi CBF menggunakan teknik TF-IDF dan *Cosine Similarity* mampu mencapai nilai *F1-Score* 0.81, yang menunjukkan performa yang tinggi [4]. Demikian pula, penelitian oleh Santoso et al. (2024) berhasil memberikan

rekomendasi film dengan akurasi dan ketepatan yang tinggi sesuai kata kunci yang dimasukan [1]. Dari penelitian yang telah disebutkan sebelumnya dan dengan bukti keberhasilan ini menjadikan *Content-Based Filtering* (CBF) sebagai landasan yang kuat untuk diimplementasikan dalam penelitian ini.

Meski penelitian sebelumnya telah membuktikan keberhasilan dari metode CBF secara teknis, masih terdapat kesenjangan antara ketersediaan metode yang teruji dengan implementasi praktisnya yang dapat diakses oleh pengguna. Banyak pengguna yang masih menghadapi permasalahan *information overload* karena terbatasnya alat bantu personalisasi yang fungsional dan terukur. Penelitian yang ada telah memberikan landasan metodologi, namun diperlukan sebuah penelitian yang berfokus pada implementasi dan evaluasi kinerja model tersebut dalam sebuah prototipe fungsional menggunakan dataset yang spesifik.

Oleh karena itu, penelitian ini dilakukan untuk menjembatani kesenjangan tersebut. Berdasarkan permasalahan *information overload* yang nyata dan potensi solusi CBF yang telah teruji secara akademis, penelitian ini bertujuan untuk mengimplementasikan model *Content-Based Filtering* menggunakan pembobotan TF-IDF dan pengukuran *Cosine Similarity* secara praktis. Tujuan utamanya adalah untuk membangun sebuah prototipe sistem rekomendasi film yang fungsional dan dapat diukur kinerjanya secara kuantitatif, sehingga dapat memberikan kontribusi praktis dalam penerapan teknologi personalisasi konten.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah penelitian ini adalah sebagai berikut :

1. Bagaimana membangun sistem rekomendasi film menggunakan metode *Content-Based Filtering* dengan menerapkan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk pembobotan fitur?
2. Seberapa besar nilai *Cosine Similarity* yang dihasilkan oleh sistem dalam menentukan tingkat kemiripan antar film berdasarkan gabungan fitur konten (soup)?
3. Seberapa besar persentase kesamaan genre serta kesamaan pemeran dan sutradara yang dicapai oleh sistem dalam memberikan daftar rekomendasi film?

## 1.3 Batasan Masalah

Agar penelitian ini lebih terfokus dan terarah, maka batasan masalah yang diterapkan adalah sebagai berikut:

1. Metode utama yang digunakan dalam penelitian ini adalah *Content-Based Filtering* (CBF) yang diimplementasikan melalui beberapa tahapan berikut:
  - a. Ekstraksi Fitur: Menggunakan TF-IDF Vectorizer untuk mengubah data teks menjadi vektor numerik. Proses ini dibatasi pada 5.000 fitur paling signifikan (*max\_features=5000*) dan menyaring kata-kata umum menggunakan *stopwords* bahasa Inggris.
  - b. Pengukuran Kemiripan: Menggunakan matrix Cosine Similarity untuk menghitung tingkat kesamaan antara film berdasarkan vektor fitur yang telah dihasilkan.
2. Batasan terkait data yang digunakan dalam penelitian ini adalah sebagai berikut:
  - a. Sumber Dataset: Dataset yang digunakan adalah movies.csv, sebuah kumpulan data film yang bersumber dari platform Kaggle.

- b. **Lingkup Data**: Data mentah yang terdiri dari 16.000 entri disaring secara signifikan menjadi 1.268 entri. Kriteria penyaringan adalah dengan hanya menyertakan film yang memiliki skor popularitas di atas 40.0.
  - c. **Fitur yang Digunakan**: Model rekomendasi dibangun menggunakan gabungan (*soup*) dari empat fitur utama: director, cast (dibatasi hanya tiga aktor pertama), genres, dan description.
  - d. **Fitur yang Diabaikan**: Beberapa Fitur sengaja tidak diikutsertakan dan dihapus.
3. **Evaluasi** : Kualitas rekomendasi dievaluasi menggunakan metrik kustom, yaitu persentase kesamaan genre dan persentase kesamaan pemeran/sutradara antara film input dengan film yang direkomendasikan.
4. **Output** : Hasil akhir penelitian ini adalah sebuah prototipe aplikasi website yang sudah diimplementasikan dengan sistem rekomendasi tersebut.



#### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan di atas, maka tujuan dari penelitian ini adalah:

1. Mengimplementasikan sistem rekomendasi film menggunakan metode *Content-Based Filtering* dengan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk pembobotan fitur.
2. Mengukur tingkat kemiripan antar film dengan menghitung nilai *Cosine Similarity* yang dihasilkan oleh sistem berdasarkan gabungan fitur konten (*soup*).
3. Mengevaluasi kualitas hasil rekomendasi dengan menghitung presentasi kesamaan genre dan persentase kesamaan pemeran atau sutradara antara film input dengan daftar film yang direkomendasikan.

#### 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat, antara lain:

1. **Manfaat Teoritis:** Penelitian ini diharapkan dapat dijadikan sebagai referensi untuk penelitian sejenis di masa mendatang di bidang sistem rekomendasi, khususnya dalam penerapan metode *Content-Based Filtering* dengan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk pembobotan dan *Cosine Similarity* untuk mengukur kemiripan.
2. **Manfaat Praktis:** Hasil penelitian ini juga dapat menjadi dasar bagi pengembang aplikasi hiburan dalam merancang fitur rekomendasi yang lebih personal dan efisien, dengan sistem rekomendasi film berbasis konten yang dapat membantu pengguna mengatasi masalah *information overload* dengan memberikan rekomendasi film yang relevan dengan preferensi mereka.

## 1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran umum mengenai isi dari setiap bab yang ada, yaitu sebagai berikut:

### BAB 1 PENDAHULUAN

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

### BAB 2 TINJAUAN PUSTAKA

Berisi ulasan penelitian terdahulu yang relevan serta penjelasan teori-teori yang mendasari penelitian ini.

### BAB 3 METODE PENELITIAN

Berisi penjelasan rinci mengenai langkah-langkah yang dilakukan dalam penelitian, mulai dari pengumpulan data hingga perancangan dan evaluasi model.

### BAB 4 HASIL DAN PEMBAHASAN

Berisi penyajian dan analisis dari seluruh hasil yang didapatkan selama penelitian, termasuk hasil evaluasi dan implementasi prototipe.

### BAB 5 PENUTUP

Berisi kesimpulan dari seluruh rangkaian penelitian serta saran untuk pengembangan di masa depan.