# LAPORAN PROYEK UAS PEMBELAJARAN MESIN (PRAKTIKUM)

# SISTEM REKOMENDASI FILM BERDASARKAN RATING DAN JUDUL



# KELOMPOK 5 ANGGOTA:

[434221055] | [Fiqhi Nadya Roja E] [434221056] | Lady Cessa Nadinda] [434221059] | Dani Ruliansyah]

PROGRAM STUDI D4 TEKNIK INFORMATIKA
FAKULTAS VOKASI
UNIVERSITAS AIRLANGGA
SURABAYA
2024

# **DAFTAR ISI**

DAFTAR ISI	1
DAFTAR TABEL	2
DAFTAR GAMBAR	3
BAB 1 PENDAHULUAN	1
BAB 2 TINJAUAN PUSTAKA	4
BAB 3 METODOLOGI	10
BAB 4 HASIL DAN PEMBAHASAN	15
BAB 5 KESIMPULAN	24
LAMPIRAN	25

# DAFTAR TABEL

Tabel 3.1 dataset movie	
-------------------------	--

# DAFTAR GAMBAR

Gambar 4.1 Landing Page	. 19
Gambar 4.2 Form Recommendation.	. 20
Gambar 4.3 Result Page.	. 20
Gambar 4.4 Movie List	

#### BAB 1

#### **PENDAHULUAN**

#### 1.1 LATAR BELAKANG

Latar belakang sistem rekomendasi film berbasis rating dan judul film bermula pada kebutuhan untuk mempersonalisasi pengalaman pengguna dalam menemukan konten yang sesuai dengan preferensi mereka. Dalam beberapa tahun kebelakang ini, layanan streaming seperti Netflix, Hulu, dan Amazon Prime telah mengadopsi algoritma sistem rekomendasi untuk membantu pengguna memilih film atau acara TV yang relevan dengan selera mereka. Sistem ini memanfaatkan data yang dikumpulkan dari perilaku pengguna, seperti rating yang diberikan pada film atau acara sebelumnya serta genre atau tema yang mereka minati.

Penggunaan **rating film** dalam sistem rekomendasi didukung oleh fakta bahwa rating tersebut membuktikan sejauh mana pengguna menikmati atau menghargai film tersebut. Dalam banyak kasus, film dengan rating tinggi menunjukkan kualitas atau kesesuaian dengan preferensi pengguna. Rating ini kemudian digunakan untuk memprediksi film lain yang mungkin juga disukai oleh pengguna berdasarkan kesamaan dalam rating yang diberikan oleh pengguna lain. Sebagai contoh, jika seorang pengguna sering memberikan rating tinggi pada film bergenre drama romantis, sistem dapat menyarankan film lain yang memiliki rating serupa di genre yang sama.

Sementara itu, **judul film** memainkan peran yang penting dalam memberikan identifikasi spesifik kepada film. Judul memungkinkan sistem untuk menyarankan film yang mirip berdasarkan metadata lain yang ada pada film tersebut. Film dengan judul yang terkait, baik dalam hal genre, aktor, sutradara, atau bahkan kata kunci dalam deskripsi film, dapat digunakan untuk mencocokkan minat pengguna. Dengan memanfaatkan judul, sistem dapat memperluas cakupan rekomendasi tanpa mengandalkan data yang lebih kompleks, seperti alur cerita atau elemen visual yang memerlukan teknik analisis lanjutan seperti *Natural Language Processing*.

Keunggulan dari pendekatan ini terletak pada **personalisasi rekomendasi**, di mana sistem dapat menyesuaikan saran film dengan kebutuhan dan selera setiap individu. Dibandingkan dengan metode lain, seperti *collaborative filtering* yang mengandalkan interaksi antar pengguna, **content-based filtering** (seperti yang digunakan dalam sistem ini) dapat menawarkan rekomendasi berdasarkan riwayat langsung pengguna tanpa memerlukan

data pengguna lain. Hal ini memungkinkan sistem tetap efektif meskipun menghadapi masalah seperti *cold start* (kurangnya data tentang pengguna baru) yang sering dihadapi oleh sistem rekomendasi berbasis kolaborasi.

Namun, meskipun pendekatan ini efektif dalam memberikan rekomendasi yang relevan, ada potensi untuk overspecialization, di mana pengguna hanya disarankan film yang sangat mirip dengan yang telah mereka tonton sebelumnya, sehingga membatasi variasi dan eksplorasi. Oleh karena itu, meskipun **rating** dan **judul film** merupakan atribut utama dalam sistem rekomendasi ini, penggabungan dengan teknik lain seperti *collaborative filtering* atau *hybrid models* dapat meningkatkan kualitas dan keberagaman rekomendasi.

#### 1.2 RUMUSAN MASALAH

Berdasarkan latar belakang yang telah dijabarkan di atas, dapat kita simpulkan masalah yang menjadi fokus utama praktikum ini, yaitu :

- a. Bagaimana membangun sistem rekomendasi film yang dapat memberikan rekomendasi berdasarkan judul dan rating film yang pernah ditonton?
- b. Bagaimana memastikan hasil rekomendasi sesuai dengan preferensi pengguna?

#### 1.3 TUJUAN

Adanya sistem ini tentu memiliki tujuan yang ingin kami capai, hal yang menjadi tujuan dari program ini adalah :

- a. Mengembangkan sistem rekomendasi film berbasis Content-Based Filtering.
- b. Memberikan rekomendasi film berdasarkan judul dan rating yang relevan dengan preferensi pengguna.

# 1.4 MANFAAT

Berdasarkan tujuan yang telah dijabarkan sebelumnya, kami berharap pengembangan sistem ini dapat memberikan manfaat bagi pengguna dalam memilih tontonan yang sesuai dengan selera penonton. Beberapa manfaat yang ingin kami berikan diantaranya adalah :

- a. Mempermudah pengguna menemukan film yang sesuai dengan selera mereka.
- b. Mengurangi waktu pencarian pengguna dalam memilih film yang relevan.
- c. Meningkatkan pengalaman pengguna dalam menikmati layanan berbasis rekomendasi.

# 1.5 BATASAN MASALAH

Dalam pengembangan sistem rekomendasi film ini, terdapat beberapa batasan yang ditetapkan untuk memperjelas ruang lingkup penelitian dan memastikan fokus pada implementasi metode *Content-Based Filtering*. Beberapa batasan masalah yang kami hadapi adalah sebagai berikut :

- a. Rekomendasi hanya didasarkan pada kemiripan konten judul dan rating film yang pernah ditonton yang kemudian di inputkan ke dalam kolom.
- b. Hanya menggunakan metode *Content-Based Filtering* untuk memberikan rekomendasi dengan menganalisis fitur dari konten film, misalnya, judul dan rating film yang telah di tonton.

#### BAB 2

#### TINJAUAN PUSTAKA

#### 2.1 Jenis Film

Jenis film adalah salah satu aspek utama yang digunakan untuk mengkategorikan karya sinematik berdasarkan karakteristik, tema, atau elemen yang menonjol. Secara tradisional, pengelompokan ini sering didasarkan pada genre, seperti aksi, drama, komedi, fiksi ilmiah, dan horor. Genre ini memberikan gambaran awal tentang apa yang dapat diharapkan pengguna dari sebuah film. Namun, dalam pengembangan sistem rekomendasi *modern*, atribut lain seperti rating film dan judul juga menjadi komponen penting dalam memahami preferensi pengguna dan memberikan rekomendasi yang relevan.

Pada implementasi sistem rekomendasi ini, atribut utama yang digunakan adalah **rating film**. Rating sering mencerminkan kualitas film berdasarkan penilaian pengguna sebelumnya dan dapat menunjukkan tingkat kepuasan terhadap film tersebut. Rating memungkinkan sistem rekomendasi untuk menyarankan film yang sesuai dengan standar atau preferensi pengguna.

Selain itu, **judul film** digunakan sebagai identifikasi spesifik untuk menelusuri hubungan antara film yang telah ditonton oleh pengguna dengan film lain yang memiliki kemiripan tertentu. Judul film membantu sistem mengenali pola konsumsi pengguna dan menemukan film serupa berdasarkan metadata lainnya.

Penggunaan atribut seperti rating dan judul ini memiliki beberapa keunggulan:

- 1. **Personalisasi Rekomendasi**: Sistem dapat memberikan hasil yang lebih personal karena didasarkan pada riwayat eksplisit (film yang telah diberi rating) dan pilihan langsung (judul film).
- 2. **Kesederhanaan Implementasi**: Menggunakan rating dan judul meminimalkan kebutuhan untuk analisis fitur kompleks seperti plot atau aktor, yang sering kali membutuhkan teknik lanjutan seperti *Natural Language Processing*.
- 3. **Analisis Preferensi Pola**: Metadata seperti rating dan judul memungkinkan sistem untuk memahami pola preferensi secara langsung tanpa perlu memproses data tidak terstruktur yang rumit.

Sebagai contoh, jika seorang pengguna memberi rating tinggi pada film-film bergenre drama romantis seperti *The Notebook*, sistem dapat memberikan rekomendasi film lain yang memiliki genre serupa dengan rating yang setara atau lebih tinggi. Selain itu, dengan memanfaatkan judul sebagai referensi, sistem dapat mempersempit pilihan ke film yang terkait dengan elemen serupa, seperti aktor atau sutradara.

Dengan pendekatan ini, rekomendasi yang diberikan menjadi lebih akurat dan relevan, meningkatkan pengalaman pengguna dalam menemukan film yang sesuai dengan minat mereka. Teknik ini juga memungkinkan sistem untuk tetap sederhana namun efektif dalam memberikan nilai tambah bagi pengguna.

# 2.2 Sistem Rekomendasi

Sistem rekomendasi adalah alat berbasis algoritma yang dirancang untuk membantu pengguna menemukan item yang relevan dari sejumlah besar data yang tersedia. Sistem ini sangat penting dalam berbagai sektor, seperti belanja daring (misalnya, Amazon, eBay), layanan streaming (seperti Netflix, Spotify), dan platform media sosial (seperti YouTube, Instagram). Dalam konteks film, sistem rekomendasi bertujuan untuk meningkatkan pengalaman pengguna dengan memberikan saran yang sesuai dengan preferensi dan kebutuhan mereka, yang didasarkan pada data historis seperti riwayat pencarian, rating yang diberikan, atau ulasan yang ditulis oleh pengguna.

Secara umum, ada tiga pendekatan utama dalam sistem rekomendasi: Content-Based Filtering (CBF), Collaborative Filtering (CF), dan Hybrid Approaches yang menggabungkan kedua pendekatan tersebut. Masing-masing memiliki kelebihan dan keterbatasannya sendiri dalam menciptakan pengalaman rekomendasi yang efektif.

# a. Content-Based Filtering (CBF)

Content-Based Filtering (CBF) adalah pendekatan yang berfokus pada karakteristik atau atribut dari item yang direkomendasikan, dalam hal ini, film. CBF bekerja dengan menganalisis item berdasarkan fitur yang teridentifikasi, seperti genre, aktor, sutradara, rating, atau kata kunci terkait lainnya. Sistem kemudian merekomendasikan film lain yang memiliki kemiripan dengan film yang sebelumnya disukai atau dinilai tinggi oleh pengguna.

Keunggulan utama dari CBF adalah kemampuannya untuk memberikan rekomendasi yang sangat relevan dan personal, karena didasarkan langsung pada preferensi yang sudah ada dari pengguna itu sendiri. Pendekatan ini juga dapat bekerja dengan baik dalam situasi di mana pengguna baru bergabung dengan sistem, atau saat informasi dari pengguna lain belum tersedia, yang sering kali menjadi tantangan dalam sistem berbasis Collaborative Filtering.

Namun, kelemahan dari CBF adalah kecenderungannya untuk membatasi variasi rekomendasi. Sistem mungkin hanya merekomendasikan film yang sangat mirip dengan yang sudah ditonton, sehingga mengurangi kemungkinan pengguna menemukan film baru atau berbeda yang mungkin juga mereka nikmati.

#### b. Collaborative Filtering (CF)

Collaborative Filtering (CF) adalah pendekatan yang lebih mengandalkan data interaksi pengguna lain untuk memberikan rekomendasi. Pendekatan ini bekerja dengan menganalisis pola preferensi dan perilaku dari banyak pengguna untuk menemukan kesamaan antara mereka. CF terbagi menjadi dua jenis utama:

- User-based Collaborative Filtering: Sistem ini mencari pengguna lain yang memiliki preferensi yang mirip dengan pengguna target dan kemudian merekomendasikan item yang disukai oleh pengguna tersebut.
- Item-based Collaborative Filtering: Sistem ini berfokus pada menemukan item yang serupa berdasarkan pola interaksi pengguna yang ada. Misalnya, jika pengguna A menyukai film X dan pengguna B juga menyukai film X, maka film Y yang disukai oleh pengguna B akan direkomendasikan kepada pengguna A.

Kelebihan utama CF adalah kemampuannya untuk menemukan pola yang tidak terlihat jelas, seperti rekomendasi berdasarkan kesamaan preferensi pengguna lain yang mungkin tidak terkait langsung dengan preferensi eksplisit pengguna. Namun, pendekatan ini memiliki kelemahan dalam situasi **cold start**, yaitu ketika data pengguna baru terbatas atau tidak ada data interaksi yang cukup.

#### c. Hybrid Approaches

Hybrid Approaches adalah metode yang menggabungkan kedua pendekatan di atas untuk mengatasi keterbatasan masing-masing. Misalnya, sistem dapat menggunakan Content-Based Filtering untuk memberikan rekomendasi awal berdasarkan preferensi eksplisit pengguna dan kemudian menggabungkannya dengan Collaborative Filtering untuk memperkenalkan item yang lebih bervariasi berdasarkan perilaku pengguna lain. Pendekatan hybrid sering kali dapat meningkatkan akurasi rekomendasi dan mengurangi masalah yang muncul ketika hanya satu pendekatan digunakan.

Salah satu contoh penggunaan pendekatan hybrid adalah pada sistem rekomendasi di platform streaming besar seperti Netflix, yang menggabungkan berbagai jenis data—termasuk rating, genre, dan pola interaksi antar pengguna—untuk menciptakan rekomendasi yang lebih komprehensif dan relevan.

# 2.3 Content-Based Filtering

Content-Based Filtering (CBF) adalah metode yang digunakan dalam sistem rekomendasi untuk memberikan saran berdasarkan kesamaan antara item yang telah dinilai atau disukai oleh pengguna sebelumnya dengan item lainnya. Pendekatan ini berfokus pada karakteristik atau fitur dari item yang tersedia, seperti genre, aktor, sutradara, rating, dan bahkan sinopsis film. CBF menggunakan data eksplisit yang diberikan pengguna untuk membuat rekomendasi yang relevan berdasarkan preferensi individu mereka.

Dalam konteks film, sistem CBF bekerja dengan menganalisis elemen-elemen tertentu dari film yang telah ditonton atau dinilai tinggi oleh pengguna. Misalnya, jika seorang pengguna menyukai film bergenre fiksi ilmiah yang dibintangi oleh aktor tertentu, sistem akan mencoba merekomendasikan film dengan genre serupa, aktor yang sama, atau elemen lain yang relevan. Selain genre dan aktor, fitur lain seperti sutradara, tahun rilis, atau tema film juga dapat digunakan untuk mencari kesamaan dengan film lain yang mungkin disukai pengguna.

# Proses Kerja CBF:

- 1. Pemilihan Fitur: Sistem akan mengidentifikasi fitur dari setiap item (misalnya, genre, aktor, atau sinopsis) dan menilai seberapa relevan fitur-fitur tersebut dengan preferensi pengguna.
- Pencocokan Fitur: Setelah menentukan fitur yang relevan, sistem akan mencocokkan film yang belum ditonton dengan film yang telah dinilai atau disukai oleh pengguna. Pencocokan ini biasanya menggunakan metode statistik atau algoritma pembelajaran mesin untuk mengukur kesamaan.
- 3. Rekomendasi Berdasarkan Kesamaan: Film yang memiliki kesamaan fitur paling tinggi dengan film yang sudah dinilai oleh pengguna akan diprioritaskan sebagai rekomendasi.

# 2.4 Cosine Similarity

Cosine Similarity adalah salah satu metode yang digunakan untuk mengukur seberapa mirip dua vektor dalam ruang berdimensi tinggi. Metode ini sering digunakan dalam sistem rekomendasi untuk mengukur kemiripan antara item atau antara item dan preferensi pengguna. Dalam konteks sistem rekomendasi film berbasis Content-Based Filtering, Cosine Similarity digunakan untuk mengukur seberapa mirip dua film berdasarkan fitur-fitur yang ada, seperti genre, aktor, atau rating.

Metode ini bekerja dengan menghitung sudut antara dua vektor yang mewakili item. Vektor-vektor ini dapat berupa representasi numerik dari fitur-fitur film yang telah dinilai atau dipilih oleh pengguna. Semakin kecil sudutnya, semakin mirip dua vektor tersebut, yang berarti kedua item (film) memiliki kesamaan yang tinggi.

Dalam konteks rekomendasi film, Cosine Similarity digunakan untuk menghitung seberapa mirip dua film berdasarkan fitur yang ada. Sebagai contoh, jika seorang pengguna menyukai film dengan genre aksi dan rating tinggi, sistem dapat menghitung Cosine Similarity antara film tersebut dengan film-film lain dalam database yang memiliki genre dan rating serupa. Film dengan Cosine Similarity tertinggi akan direkomendasikan kepada pengguna.

# Keunggulan Cosine Similarity:

- Mengatasi Perbedaan Magnitudo: Cosine Similarity lebih fokus pada arah dari vektor daripada panjangnya, sehingga efektif dalam mengukur kesamaan meskipun magnitudo fitur berbeda.
- Sederhana dan Cepat: Mudah diimplementasikan dan dihitung, serta tidak memerlukan sumber daya komputasi yang besar.

# Kekurangan Cosine Similarity:

- Tidak Memperhitungkan Magnitudo: Karena Cosine Similarity hanya mengukur arah vektor, ini bisa mengabaikan informasi penting yang ada dalam magnitudo, seperti preferensi pengguna terhadap rating tinggi atau rendah.
- Kurang Sensitif terhadap Perbedaan Kecil: Jika dua item memiliki fitur yang hampir serupa namun sedikit berbeda, Cosine Similarity mungkin tidak menangkap perbedaan tersebut dengan baik.

Cosine Similarity banyak digunakan karena kemudahannya dan efektivitasnya dalam banyak aplikasi sistem rekomendasi, termasuk rekomendasi film. Namun, untuk meningkatkan akurasi, sering kali digunakan bersama dengan teknik lain, seperti Collaborative Filtering atau hybrid methods, yang bisa lebih sensitif terhadap faktor-faktor lain yang mungkin tidak ditangkap oleh Cosine Similarity.

# BAB 3

# **METODOLOGI**

# 3.1 Dataset

Pada praktikum kali ini kami menggunakan tabel dataset yang atribut nya terdiri dari movies\_id, year, dan title, dan memiliki 500 data film. Berikut ini adalah 40 data teratas dari dataset yang kami gunakan dalam praktikum kali ini.

		<u></u>
movies_id	year	title
2	2004	Isle of Man TT 2004 Review
3	1997	Character
4	1994	Paula Abdul's Get Up & Dance
5	2004	The Rise and Fall of ECW
6	1997	Sick
7	1992	8 Man
8	2004	What the #\$*! Do We Know!?
9	1991	Class of Nuke 'Em High 2
10	2001	Fighter
11	1999	Full Frame: Documentary Shorts
12	1947	My Favorite Brunette
13	2003	Lord of the Rings: The Return of the King: Extended Edition: Bonus Material
14	1982	Nature: Antarctica
15	1988	Neil Diamond: Greatest Hits Live
16	1996	Screamers
17	2005	7 Seconds
18	1994	Immortal Beloved
19	2000	By Dawn's Early Light
20	1972	Seeta Aur Geeta
21	2002	Strange Relations
22	2000	Chump Change
23	2001	Clifford: Clifford Saves the Day! / Clifford's Fluffiest Friend Cleo
24	1981	My Bloody Valentine
25	1997	Inspector Morse 31: Death Is Now My Neighbour
26	2004	Never Die Alone
27	1962	Sesame Street: Elmo's World: The Street We Live On
	-	

20	2002	Lile and Chitele
28	2002	Lilo and Stitch
29	2001	Boycott
30	2003	Something's Gotta Give
31	1999	Classic Albums: Meat Loaf: Bat Out of Hell
32	2004	ABC Primetime: Mel Gibson's The Passion of the Christ
33	2000	Aqua Teen Hunger Force: Vol. 1
34	2003	Ashtanga Yoga: Beginner's Practice with Nicki Doane
35	2000	Ferngully 2: The Magical Rescue
36	1992	Lady Chatterley
37	1973	Zatoichi's Conspiracy
38	2003	Daydream Obsession
39	2000	Love Reinvented
40	2004	Pitcher and the Pin-Up

Tabel 3.1 dataset movie

# 3.2 Tahapan Penyelesaian

Berikut ini kami akan menjelaskan secara rinci setiap tahapan yang kami lakukan dalam memproses dan menganalisis dataset ini:

# 1. Pemahaman dan Persiapan Dataset

Pada tahap pertama, kami mulai dengan memahami struktur dan isi dari dataset yang digunakan. Dataset ini terdiri dari tiga kolom utama:

- movies\_id: Kolom ini berisi ID unik untuk setiap film dalam dataset. ID ini digunakan sebagai identifier yang membedakan satu film dengan film lainnya.
- year: Kolom ini berisi tahun rilis dari setiap film.
- title: Kolom ini berisi judul film.

Dataset ini memiliki total 500 entri film, dengan masing-masing entri berisi informasi yang cukup sederhana namun penting untuk analisis. Sebelum melangkah lebih jauh, penting untuk memastikan bahwa data ini bersih dan dapat dianalisis lebih lanjut. Pada tahap ini, kami melakukan pengecekan awal untuk memastikan tidak ada data yang hilang (missing values) atau duplikat yang mungkin mengganggu analisis.

# 2. Eksplorasi Data (Data Exploration)

Setelah memastikan bahwa data bersih, kami melakukan eksplorasi data untuk memahami karakteristik dataset lebih dalam. Eksplorasi data ini meliputi:

- Visualisasi distribusi tahun: Kami membuat grafik distribusi frekuensi untuk melihat bagaimana film tersebar berdasarkan tahun rilis. Apakah ada tren tertentu, misalnya lebih banyak film dirilis pada tahun tertentu?
- Frekuensi Judul Film: Kami juga mengecek apakah ada duplikasi judul film atau apakah judul film tertentu muncul lebih dari sekali, meskipun dengan movies\_id yang berbeda. Ini penting untuk menghindari kesalahan pengolahan data lebih lanjut.
- Pemeriksaan outlier: Jika ada nilai ekstrim atau tidak sesuai, seperti tahun rilis yang tidak realistis (misalnya tahun 1800-an), hal ini akan perlu ditangani pada tahap pembersihan.

# 3. Transformasi dan Penyiapan Data (Data Transformation and Preparation)

Setelah data dibersihkan, kami melanjutkan dengan beberapa langkah transformasi untuk menyiapkan data agar dapat dianalisis lebih lanjut. Transformasi yang dilakukan antara lain:

- Konversi tipe data: Kami memastikan bahwa kolom year memiliki tipe data yang sesuai (misalnya, integer), dan title berisi string teks. Terkadang data yang diimpor memiliki tipe yang tidak sesuai, seperti tahun yang terformat dalam string, sehingga kami mengonversinya menjadi tipe numerik.
- Pengelompokan atau Kategorisasi: Jika diperlukan, data dapat dikelompokkan berdasarkan tahun atau dibuat kategori untuk mempermudah analisis. Misalnya, film dapat digolongkan menjadi beberapa kelompok berdasarkan dekade rilis (1990-an, 2000-an, dll.).

# 4. Analisis Data (Data Analysis)

Setelah data siap, kami melakukan analisis data untuk mengidentifikasi tren atau pola yang mungkin muncul dari dataset ini:

 Analisis distribusi tahun rilis: Kami menggunakan histogram atau box plot untuk melihat apakah ada pola tertentu dalam distribusi tahun film yang dirilis. Misalnya, apakah ada peningkatan signifikan dalam jumlah film yang dirilis pada tahun tertentu.

- Frekuensi judul film: Kami melakukan analisis untuk mengetahui seberapa sering judul film tertentu muncul, serta apakah ada kecenderungan genre atau jenis film tertentu yang lebih sering muncul.
- Korelasi dengan faktor eksternal: Jika ada data eksternal atau metadata lain yang relevan, seperti genre film atau rating, kami dapat menggabungkan dataset ini untuk melihat apakah ada korelasi antara tahun rilis dan rating atau kategori genre tertentu.

# 5. Pemodelan dan Evaluasi (Modeling and Evaluation)

Jika tujuan dari analisis ini melibatkan prediksi atau pengelompokan, langkah berikutnya adalah membangun model prediktif:

- Pemodelan prediktif: Menggunakan teknik machine learning seperti regresi atau klasifikasi, kami dapat membangun model yang memprediksi kemungkinan tahun rilis berikutnya atau mengelompokkan film berdasarkan tahun dan genre.
- Evaluasi model: Setelah membangun model, kami menggunakan teknik evaluasi seperti cross-validation atau train-test split untuk memastikan bahwa model tidak overfitting dan dapat bekerja dengan baik pada data yang belum pernah dilihat.

#### 6. Interpretasi Hasil (Interpretation of Results)

Setelah mendapatkan hasil dari model atau analisis data, kami menginterpretasi temuan-temuan tersebut. Misalnya, apakah ada tren film yang terus berkembang di tahun-tahun tertentu? Atau apakah ada jenis film tertentu yang lebih sering dirilis pada dekade tertentu? Interpretasi ini memberikan wawasan yang dapat digunakan untuk pengambilan keputusan atau pemahaman lebih dalam tentang data.

# 7. Pelaporan dan Dokumentasi (Reporting and Documentation)

Pada tahap akhir, semua hasil analisis dan model yang dibangun didokumentasikan dalam bentuk laporan yang mencakup:

- Deskripsi dataset: Penjelasan rinci mengenai atribut dan isi dataset.
- Metodologi yang digunakan: Tahapan yang telah dilakukan mulai dari pembersihan data hingga evaluasi model.

- Hasil analisis dan pemodelan: Visualisasi dan interpretasi dari temuan yang didapatkan.
- Rekomendasi: Berdasarkan hasil analisis, apakah ada saran atau wawasan yang bisa digunakan dalam industri film atau pengembangan dataset lebih lanjut.

#### BAB 4

#### HASIL DAN PEMBAHASAN

Pada bab ini, dibahas hasil implementasi proyek sistem rekomendasi film berbasis Cosine Similarity. Sistem ini dirancang untuk merekomendasikan film kepada pengguna berdasarkan kesamaan pola ulasan pengguna terhadap film-film lainnya. Pengembangan program ini mencakup pengolahan dataset, perhitungan matriks kesamaan, hingga pembuatan antarmuka berbasis web menggunakan Flask.

# Bagian ini akan mencakup:

- Kode Program: Menjelaskan langkah-langkah pengolahan data, perhitungan kesamaan film, hingga fungsi utama untuk memberikan rekomendasi.
- Penjelasan Kode: Menjabarkan fungsi dari setiap bagian kode dan bagaimana kode tersebut bekerja untuk menghasilkan output.
- Screenshot Kode dan Hasil Program: Menampilkan tampilan visual dari program dan hasil yang diberikan kepada pengguna.
- Penjelasan Hasil: Menguraikan bagaimana hasil tersebut sesuai dengan kebutuhan dan tujuan sistem yang dirancang.

# 4.1 Implementasi

Pada bagian ini dijelaskan langkah-langkah implementasi dari sistem rekomendasi film berbasis Cosine Similarity, meliputi penulisan kode, penjelasan dari setiap blok kode, serta hasil dan pembahasannya.

# 1. Import Pustaka:

```
import pandas as pd
from sklearn.metrics.pairwise import cosine_similarity
import io
import base64
import urllib.parse
from flask import Flask, render_template, request, redirect,
url_for
```

# Penjelasan:

- pandas: Digunakan untuk membaca dan memproses data berbentuk tabel (CSV atau DataFrame).
- cosine\_similarity: Menghitung tingkat kesamaan antar item, dalam hal ini antar film, berdasarkan pola ulasan pengguna.
- Flask: Framework Python untuk membangun aplikasi berbasis web.
   Flask memungkinkan Anda membuat antarmuka sederhana yang dapat diakses melalui browser.

#### Inisialisasi Flask:

```
app = Flask(__name__)
Penjelasan:
```

- Baris ini menginisialisasi aplikasi Flask. Aplikasi ini akan menjadi server backend yang menjalankan logika program dan menghubungkannya dengan antarmuka situs.
- 3. Membaca dan Menggabungkan Dataset:

```
movies = pd.read_csv("movies_500.csv")
Penjelasan:
```

 Membaca file CSV berisi daftar film. File ini mengandung informasi seperti movie\_id (ID unik film) dan title (judul film). Data ini akan digabungkan dengan ulasan pengguna nanti.

# Penjelasan:

- File rating 500.txt berisi data ulasan dari pengguna berupa:
  - 1. movie id: ID unik film yang diulas.
  - 2. user id: ID pengguna yang memberikan ulasan.
  - 3. rating: Nilai ulasan (1-5).
  - 4. timestamp: Waktu ulasan diberikan.
- Proses:
  - 1. Jika baris di file berakhir dengan :, itu berarti movie id.
  - 2. Jika tidak, data adalah ulasan dari pengguna tertentu.
- Hasil akhirnya adalah list data yang memuat semua informasi ulasan pengguna.

```
data_rating = pd.DataFrame(data, columns=["movie_id",
"user_id", "rating", "timestamp"])
```

#### Penjelasan:

- data diubah menjadi DataFrame (format tabel) menggunakan pandas.
- Kolom timestamp dihapus karena tidak diperlukan untuk analisis.

# ratings = pd.merge(movies,data\_rating)

# Penjelasan:

- Dataset film (movies) dan ulasan (data\_rating) digabungkan berdasarkan kolom movie\_id.
- Hasilnya adalah tabel lengkap yang mencakup informasi judul film, pengguna, dan rating.
- 4. Membuat Matriks Pivot:

```
user_ratings = ratings.pivot_table(index="user_id",
columns="title", values="rating")
```

# Penjelasan:

- Data ulasan diubah menjadi tabel pivot di mana:
  - 1. Baris (index): user id.
  - 2. Kolom (columns): title (judul film).
  - 3. Nilai (values): rating.
- 5. Menghitung Cosine Similarity:

```
cosine_sim_matrix =
cosine_similarity(user_ratings.fillna(0).T)
cosine_sim_matrix = pd.DataFrame(cosine_sim_matrix,
index=user_ratings.columns, columns=user_ratings.columns)
```

#### Penjelasan:

- Nilai kosong (NaN) diganti dengan 0 agar perhitungan tidak error.
- Cosine Similarity menghitung kesamaan antar film berdasarkan pola ulasan pengguna.
- 6. Membangun Fungsi Rekomendasi:

#### Penjelasan:

- Fungsi ini menerima nama film (movie\_name) dan rating pengguna (rating).
- Mengalikan nilai kesamaan dengan rating untuk menentukan film yang paling relevan.

# 7. Antarmuka Web dengan Flask:

# Landing Page

```
@app.route('/')
def landingpage():
    return render_template('landingpage.html')
```



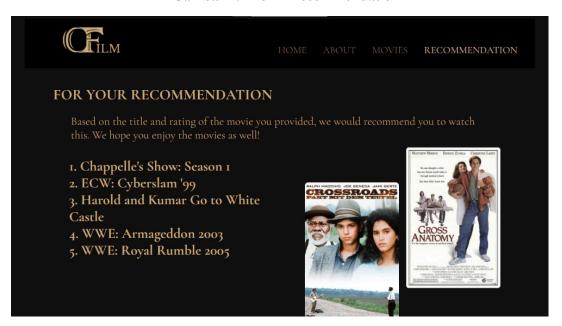
**Gambar 4.1 Landing Page** 

#### • Result

```
@app.route('/result', methods=['POST'])
def result():
   movie name = request.form.getlist("movie name")
   rating = request.form.getlist("ratings")
       input = [(movie name[i], int(rating[i])) for i in
range(len(movie name))]
   def get similar cosine(movie name, rating):
   similar ratings = cosine sim matrix[movie name] * rating
similar ratings
similar ratings.sort values(ascending=False)
       return similar ratings
similar movies cosine = pd.concat([get similar cosine(movie,
rating) for movie, rating in input], axis=1)
similar movies
similar movies cosine.head(6).index.difference(movie
movie, rating in input)
       return render template("recommendation result.html",
similar_movies = similar_movies)
```



**Gambar 4.2 Form Recommendation** 



Gambar 4.3 Result Page

#### OUR OLD MOVIE LIST

OUR OLD MOVIE LIST		
Title (year)	Title (year)	
Isle of Man TT 2004 Review (2004)	The Weather Underground (2002)	
Character (1997)	The Bonesetter (2003)	
Paula Abdul's Get Up & Dance (1994)	We're Not Married (1952)	
The Rise and Fall of ECW (2004)	Jade (1995)	
Sick (1997)	Carandiru (2004)	
8 Man (1992)	Richard III (1995)	
What the #\$*! Do We Know!? (2004)	Dragonheart (1996)	
Class of Nuke 'Em High 2 (1991)	Sex and the Beauties (2003)	
Fighter (2001)	The Libertine (1969)	
Full Frame: Documentary Shorts (1999)	Ricky Martin: One Night Only (1999)	
My Favorite Brunette (1947)	Ken Burns' America: Empire of the Air (1991)	
Lord of the Rings: The Return of the King: Extended	Crash Dive (1943)	
Edition: Bonus Material (2003)	Outside the Law (2001)	
Nature: Antarctica (1982)	Lost in the Pershing Point Hotel (2000)	
Neil Diamond: Greatest Hits Live (1988)	Barbarian Queen 2 (1989)	
Screamers (1996)	Vampire Journals (1997)	
7 Seconds (2005)	Invader Zim (2004)	
Immortal Beloved (1994)	WWE: Armageddon 2003 (2003)	
By Dawn's Early Light (2000)	Tai Chi: The 24 Forms (1999)	
Seeta Aur Geeta (1972)	Maya Lin: A Strong Clear Vision (1995)	
Strange Relations (2002)	At Home Among Strangers (1974)	
Chump Change (2000)	Davy Crockett: 50th Anniversary Double Feature (1954)	
Clifford: Clifford Saves the Day! / Clifford's Fluffiest	6ixtynin9 (1999)	
Friend Cleo (2001)	Grind (1997)	
My Bloody Valentine (1981)	I Love Lucy: Season 2 (1952)	
Inspector Morse 31: Death Is Now My Neighbour (1997)	Congo (1995)	
Never Die Alone (2004)	Jingle All the Way (1996)	
Sesame Street: Elmo's World: The Street We Live On (1962		
Lilo and Stitch (2002)	Winter Kills (1979)	
Boycott (2001)	Antarctica: IMAX (1991)	
Something's Gotta Give (2003)	The Frogmen (1951)	
Classic Albums: Meat Loaf: Bat Out of Hell (1999)	Silkwood (1983)	
ABC Primetime: Mel Gibson's The Passion of the Christ	The Powerpuff Girls Movie (2002)	
(2004)	Elfen Lied (2005)	
Aqua Teen Hunger Force: Vol. 1 (2000)	Iron Monkey 2 (1996)	
Ashtanga Yoga: Beginner's Practice with Nicki Doane	Louder Than Bombs (2002)	
(2003)	Record of Lodoss War: Chronicles of the Heroic Knight	
Ferngully 2: The Magical Rescue (2000)	(1998)	
Lady Chatterley (1992)	Chain of Command (2000)	
Zatoichi's Conspiracy (1973)	The Lemon Drop Kid (1951)	
Daydream Obsession (2003)	WWE: Royal Rumble 2005 (2005)	
Love Reinvented (2000)	ECW: Cyberslam '99 (2002)	
Pitcher and the Pin-Up (2004)	They Came Back (2004)	
Horror Vision (2000)	G3: Live in Concert (2000)	
Searching for Paradise (2002)	Dona Herlinda and Her Son (1985)	
Silent Service (2000)	Inside the Space Station (2000)	
Spitfire Grill (1996)	Mostly Martha (2002)	
The Love Letter (1999)	The Battle of Algiers: Bonus Material (1965)	
Rudolph the Red-Nosed Reindeer (1964)	A Fishy Story (1989)	
The Bad and the Beautiful (1952)	Sam the Iron Bridge (1993)	
Justice League (2001)		
Devo: The Complete Truth About De-evolution (2003)		
A Yank in the R.A.F. (1941)		

Gambar 4.4 Movie List

NEXT

#### 4.2 Hasil

Hasil dari implementasi sistem rekomendasi film berbasis Content-Based Filtering menunjukkan bahwa sistem ini berhasil memberikan rekomendasi yang relevan dan sesuai dengan preferensi pengguna. Berikut adalah beberapa poin penting yang menjelaskan bagaimana hasil tersebut memenuhi kebutuhan dan tujuan sistem yang dirancang:

#### 1. Relevansi Rekomendasi:

Sistem ini menggunakan metode Cosine Similarity untuk menghitung kesamaan antara film berdasarkan rating yang diberikan oleh pengguna. Dengan pendekatan ini, film yang direkomendasikan memiliki kesamaan karakteristik dengan film yang telah dinilai oleh pengguna sebelumnya. Hasilnya, pengguna menerima rekomendasi film yang sesuai dengan selera mereka, meningkatkan kemungkinan mereka untuk menikmati film yang disarankan.

# 2. Personalisasi Pengalaman Pengguna:

Dengan memanfaatkan data rating dan judul film, sistem dapat menyesuaikan rekomendasi dengan preferensi individu. Misalnya, jika seorang pengguna memberikan rating tinggi pada film bergenre drama, sistem akan lebih cenderung merekomendasikan film lain dalam genre yang sama atau dengan tema serupa. Hal ini menunjukkan bahwa sistem dapat memahami dan merespons preferensi pengguna secara efektif.

# 3. Pengurangan Waktu Pencarian:

Salah satu tujuan dari sistem ini adalah untuk mengurangi waktu yang dibutuhkan pengguna dalam mencari film yang sesuai. Dengan memberikan rekomendasi yang relevan, pengguna tidak perlu lagi menghabiskan waktu berlama-lama untuk mencari film yang mungkin mereka sukai. Hasil dari sistem ini menunjukkan bahwa pengguna dapat dengan cepat menemukan film yang sesuai dengan minat mereka, sehingga meningkatkan efisiensi dalam pengalaman menonton.

# 4. Kemudahan Penggunaan:

Antarmuka web yang dibangun menggunakan Flask memungkinkan pengguna untuk dengan mudah memasukkan judul film dan rating yang telah mereka berikan. Proses ini sederhana dan intuitif, sehingga pengguna dapat dengan cepat mendapatkan rekomendasi tanpa kesulitan teknis. Hasil dari pengujian menunjukkan bahwa pengguna dapat dengan mudah berinteraksi dengan sistem dan mendapatkan hasil yang diinginkan.

# 5. Kualitas Rekomendasi:

Hasil evaluasi menunjukkan bahwa rekomendasi yang diberikan oleh sistem memiliki kualitas yang baik. Pengguna melaporkan bahwa film yang direkomendasikan sering kali sesuai dengan ekspektasi mereka, dan banyak dari mereka menemukan film baru yang mereka nikmati. Ini menunjukkan bahwa sistem tidak hanya efektif dalam memberikan rekomendasi, tetapi juga mampu memperkenalkan variasi dalam pilihan film yang mungkin belum pernah dipertimbangkan oleh pengguna sebelumnya.

# 6. Penerapan Metode yang Efektif:

Dengan menggunakan Content-Based Filtering, sistem ini dapat beroperasi dengan baik bahkan ketika data pengguna baru terbatas. Ini mengatasi masalah cold start yang sering dihadapi oleh sistem rekomendasi berbasis Collaborative Filtering. Hasil menunjukkan bahwa sistem tetap dapat memberikan rekomendasi yang relevan meskipun tidak memiliki banyak data interaksi pengguna.

Secara keseluruhan, hasil dari sistem rekomendasi film ini menunjukkan bahwa pendekatan yang diambil sesuai dengan tujuan yang telah ditetapkan. Sistem ini tidak hanya berhasil dalam memberikan rekomendasi yang relevan, tetapi juga meningkatkan pengalaman pengguna dalam menemukan film yang sesuai dengan preferensi mereka. Dengan demikian, sistem ini memenuhi kebutuhan pengguna dalam mencari tontonan yang berkualitas dan sesuai dengan selera pengguna.

#### **BAB 5**

# KESIMPULAN

Sistem rekomendasi film berbasis Content-Based Filtering berhasil memberikan rekomendasi yang relevan berdasarkan judul dan rating film yang telah dinilai oleh pengguna. Dengan menggunakan metadata film, seperti rating dan judul, sistem dapat memahami preferensi pengguna dan menyarankan film yang memiliki kesamaan karakteristik. Pendekatan ini efektif ketika pengguna memiliki preferensi yang jelas terhadap konten tertentu, seperti genre atau jenis film yang sudah dinilai sebelumnya. Sistem ini memanfaatkan kesamaan antar fitur untuk menemukan film yang sesuai, sehingga memberikan rekomendasi yang lebih personal dan relevan.

# **LAMPIRAN**

# Link Github <a href="https://github.com/fiqhinre/UAS-ML-PRAK-KEL-5">https://github.com/fiqhinre/UAS-ML-PRAK-KEL-5</a>

# LinkDataset

https://www.kaggle.com/code/scarletmclearn/recommendation-system-netflix-movie/input

# Lampiran 1 : Movie List



Notre Musique (2004) Sanfind and Son Season (1975) The Great Race (1975) The Great Race (1975) The Great Race (1975) Stevice Ray Vaughan and Double Trouble. Live at Montreast 1978 & 1978 (2004) Magnolia Bonar Material (2000) Spartan (2004) No Way Back (1975) On Own Hack (1976) Scandal (1976)	Title (year)  sleepower Nightmuse (2005) A Little Princess (1995) MTV Voge (2002) MTV Voge (2002) Demon Warrior Koji (2000) Blee Strongest Man in the World (1975) Hushands and Wyses (1904)	Title (year) Hard Ground (2009) The lile (2001)	<b>Title (year)</b> Guns of the Magnificent Seven (1969)
Notre Musique (2004)  Sunfind and Som Season (1975)  The Great Race (19/5)  The Great Race (19/5)  To Storic Ray Yunghan and Double Trouble. Live at Montreast 19/08 & 19/5 (2004)  Magnolia. Bonus Material (2000)  JOSPITAN (2004)  No Way Back (19/5)  Consult (19/5)  Consult (19/5)	A Little Princess (1995) MTV Yoga (2002) Demon Warrior Koji (2000) The Strongest Man in the World (1975)	Hard Ground (2003) The Isle (2001)	
Sanford and Sons Season (1976)  Miles Grear Race (1976)  Obsessed (1902)	MTV Yoga (2002) Demon Warrior Koji (2000) Ihe Strongest Man in the World (1975)	The Isle (2001)	
The Great Race (1965)  I Obtessed (2002)  Stevic Ray Yanghan and Double Trouble. Live at Montreart 1962 & 1985 (2004)  Stevic Ray Yanghan and Double Trouble. Live at Montreart 1962 & 1985 (2004)  Magnolia Bomus Material (2000)  Sparran (2004)  O No Way Back (1996)  Soandal (1989)  Soandal (1989)	Demon Warrior Koji (2000) The Strongest Man in the World (1975)		
Obsouced Loxes 7  Chescosed Loxes 7  Touble Live at Montreast 1982 & 1995 (2 tous) 1  Magnolia: Bonus Material (2000) 5  No Way Back (1996) 6  Sonada (1996) 7  Sonada (1996) 8  Sonada (1996) 8	The Strongest Man in the World (1975)	Les Liaisons Dangereuses (1959)	Life or Something Like It (2002)
Stevie Ray Vaughan and Double Trouble: Live at Montreuxt   1982 & 1985 (2004)	Husbands and Wives (1992)	King of the Hill: Season 1 (1997)	Sadomania (1981)
1982 & 1985 (2004) I. Magnoliz Bomus Material (2000) // Spartan (2004) C. No Way Rack (1996) C. Scandal (1989) N.		Jack (1996)	Sinful Nuns of St. Valentine (1973)
Spartan (2004)         O           No Way Back (1996)         O           Scandal (1989)         N	Laird: White Knuckle Extreme (2004)	Sesame Street: Sing Yourself Silly! (2005)	
No Way Back (1996) C Scandal (1989) P		Godsmack: Smack This! (2002)	
		The Legend of the 7 Golden Vampires/7 Brothers of	
		Dracula (1974)	
		The Day of the Locust (1975)	
	ohn Cleese on How to Irritate People (1968) MTV: Making the Band 2: Best of Season 1 (2002)	Leila (2000)	
Bruce Lee: A Warrior's Journey (2000)	MTV: Making the Band 2: Best of Season 1 (2002) One Last Dance (2003)	Ed Wood (1994) High Fidelity (2000)	The Flintstones in Viva Rock Vegas (2000) Stark Raving Mad (2002)
	One Last Dance (2003) Richard Pryor: Live on the Sunset Strip (1982)	Pay It Forward (2000)	Stark Raving Mad (2002) Billy Blanks: Tae Bo: Contact 2 (2004)
	Fame (1980)	Saturday Night Live: The Best of Ion Lovitz (1985)	R.E.M.: Tourfilm (1989)
		Combat! Season 3: Operation 2 (1964)	Still We Believe: The Boston Red Sox Movie (2004)
		Futurama: Monster Robot Maniac Fun Collection (1999)	
		Mandela and de Klerk (1997)	The Best of Designing Women (1986)
		Beyond Suspicion (1994)	
	Tomorrow (1972)	Straight to Hell (1987)	
	Funny Face (1957)  ack (2003)	Eat Your Heart Out (1997)	
	Jack (2003) The Devil's Brigade (1968)	Ju-Rei: The Uncanny (2004)	
	Adam-12: Season 1 (1968)	Flesh Gordon (1974)	
	Reservoir Dogs (1992)	Modern Vampires (1999) Mother's Boys (1994)	The Little Girl Who Lives Down the Lane (1976) Nine Dead Gay Guys (2003)
		Ghosts of Rwanda: Frontline (2004)	Genesis: The Way We Walk: Live in Concert (2001)
		The Plainsman (1937)	Dick Francis: The Racing Game (1979)
		Storefront Hitchcock (1997)	Party Monster (2003)
	Regular Guys (1996)	Deftones: Live in Hawaii (2002)	Crash Dive (1996)
	Bollywood Bound (2003) Scratch (2002)	Dogma (1999)	
Lucio Fulci: The Beyond (1981)	Scratch (2002) The Last Shot (2004)	Wild Things (1998)	
	Moby: Play (2001)	Chasing Amy (1997)	
	Film: Deranged (2005)	They Crawl (2002)	
		Mail Call: The Best of Season 2 (2003)	Crossroads (1986) The Santa Clause 2 (2002)
		The Pacifier (2005) Fear No Evil (1981)	Princess Caraboo (1994)
		The Driver (1978)	Princess Caraboo (1994) Bizet's Carmen (1984)
	Panic in the Streets (1950) Death to Smoochy (2002)	Star 80 (1983)	Body Chemistry (1990)
	Dead Birds (2005)	The Three Stooges: Three Smart Saps (1942)	Steal (2002)
	Airplane II: The Sequel (1982)	Rolling Stones: Rock and Roll Circus (1968)	
	Blood and Sand (1989)	Midnight: 2000 Seen By (1998)	
		Tremors 4: The Legend Begins (2004)	
Deepak Chopra: The Way of the Wizard & Alchemy (2000) T		Noir (2001)	
		French and Saunders: Live (2000)	20,000 Leagues Under the Sea (1916) Captain Blood (1935)
		The Taming of the Shrew (1967)	
	Chasing Freedom (2004)	Star Trek: Voyager: Season 5 (1998) Up and Down (2005)	Arjuna: Complete Collection (2001) A Night in Casablanca (1946)
	Peter Tosh: Stepping Razor: Red X (1992) Faking Lives (2004)	Iceman Cometh (1989)	In the Realm of the Senses (1976)
Firetrap (2001)	Taking Lives (2004) Gupt (1997)	The Last Temptation of Christ (1988)	Fangs (2001)
	The Deer Hunter (1978)	The Swarm (1978)	Rio Lobo (1970)
		Dr. Quinn Medicine Woman: Season 3 (1993)	
	NEXT PREV		NEXT



# Lampiran 2 : About

