

**TUGAS SISTEM PENDUKUNG KEPUTUSAN
“SISTEM REKOMENDASI FILM BERDASARKAN USIA”**



**Disusun oleh:
Agung Bahtiar (362258302093)**

**PROGRAM STUDI TEKNOLOGI REKAYASA PERANGKAT LUNAK
JURUSAN TEKNIK INFORMATIKA
POLITEKNIK NEGERI BANYUWANGI
2024**

Merancang Sistem Pendukung Keputusan Rekomendasi Film Berdasarkan Usia

Untuk menciptakan sistem cerdas yang andal dalam menyarankan film, kita perlu mengumpulkan beberapa informasi kunci sebagai dasar sistem tersebut. Informasi penting yang diperlukan mencakup:

Nama film: Penanda khusus untuk setiap judul dalam kumpulan data.

Skor: Nilai rata-rata yang diberikan oleh pemirsa atau pengulas film.

Kategori: Klasifikasi atau tipe film yang mencerminkan isi dan tema utamanya.

Jumlah pemberi nilai: Total penilaian dari penonton, menunjukkan tingkat kepopuleran dan keandalan skor.

Keempat unsur data ini akan menjadi komponen utama dalam metode rekomendasinya, memungkinkan sistem mengkaji kesukaan pengguna dan memberikan saran film yang sesuai dan tepat.

Untuk menjamin ketepatan dan keefektifan sistem cerdas rekomendasi film ini, kami akan menjalankan dua langkah dalam proses pengolahan informasi, yaitu **Pemahaman Data** dan **Penyiapan Data**.

Pemahaman Data

Pengumpulan Informasi

Untuk Merancang Sistem Pendukung Keputusan Rekomendasi Film Berdasarkan Usia, kami memanfaatkan dataset film dari Kaggle. Sumber ini menyajikan informasi lengkap dan dapat diandalkan mengenai film-film unggulan, termasuk judul, skor, kategori, dan detail penting lainnya, yang sangat diperlukan untuk mengembangkan sistem rekomendasi yang tepat dan efisien.

IMDB Movies Dataset

Top 1000 Movies by IMDB Rating

[Data Card](#) [Code \(173\)](#) [Discussion \(4\)](#) [Suggestions \(0\)](#)

About Dataset

Context

IMDB Dataset of top 1000 movies and tv shows.

You can find the **EDA Process** on - <https://www.kaggle.com/harshitshankhdhar/eda-on-imdb-movies-dataset>

Please consider **UPVOTE** if you found it useful.

Content

Data:-

- **Poster_Link** - Link of the poster that imdb using
- **Series_Title** = Name of the movie
- **Released_Year** - Year at which that movie released
- **Certificate** - Certificate earned by that movie
- **Runtime** - Total runtime of the movie
- **Genre** - Genre of the movie
- **IMDB_Rating** - Rating of the movie at IMDB site
- **Overview** - mini story/ summary
- **Meta_score** - Score earned by the movie
- **Director** - Name of the Director

Pengenalan Pola Data

Dalam proses pengkajian data, kami akan menggunakan alat Pandas dan NumPy dari Python untuk mengenali dan menjelajahi pola-pola dalam kumpulan data.

Mengimpor Alat Analisis

Kami menggunakan dua alat Python utama: Pandas untuk mengelola dan menganalisis data terstruktur, serta NumPy untuk perhitungan numerik yang efisien. Kedua alat ini menjadi dasar dalam pengolahan dan analisis statistik untuk sistem penyaran film kami.

Memulai Pengkajian Data

Kami menggunakan Pandas untuk membaca kumpulan data dari file CSV yang berisi informasi film IMDb. Data ini kemudian disimpan dalam bentuk DataFrame untuk analisis lebih lanjut.

```

import pandas as pd
import numpy as np

path = "/dataset/imdb_top_1000.csv"
df = pd.read_csv(path);
df.head()

```

	Poster_Link	Series_Title	Released_Year	Certificate	Runtime	Genre	IMDB_Rating	Overview	Meta_score	Director	Star1	Star2	Star3	Star4	No_of_Votes
0	amazon.com/images/M/MV5BMDFKYT...	The Shawshank Redemption	1994	A	142 min	Drama	9.3	Two imprisoned men bond over a number of years...	80.0	Frank Darabont	Tim Robbins	Morgan Freeman	Bob Gunton	William Sadler	2343110
1	amazon.com/images/M/MV5BMDZyNj...	The Godfather	1972	A	175 min	Crime, Drama	9.2	An organized crime dynasty's aging patriarch...	100.0	Francis Ford Coppola	Marlon Brando	Al Pacino	James Caan	Diane Keaton	1620367
2	amazon.com/images/M/MV5BMTMxNT...	The Dark Knight	2008	UA	152 min	Action, Crime, Drama	9.0	When the menace known as the Joker wreaks havoc...	84.0	Christopher Nolan	Christian Bale	Heath Ledger	Aaron Eckhart	Michael Caine	2303232
3	amazon.com/images/M/MV5BMDMwMG...	The Godfather: Part II	1974	A	202 min	Crime, Drama	9.0	The early life and career of Vito Corleone in...	90.0	Francis Ford Coppola	Al Pacino	Robert De Niro	Robert Duvall	Diane Keaton	1129952
4	amazon.com/images/M/MV5BMDU4N2...	12 Angry Men	1957	U	96 min	Crime, Drama	9.0	A jury holdout attempts to prevent a miscarria...	96.0	Sidney Lumet	Henry Fonda	Lee J. Cobb	Martin Balsam	John Fiedler	689845

Memeriksa Jenis Data

Kami melihat jenis data setiap kolom untuk memahami struktur data dan merencanakan strategi pengolahan yang tepat.

```

print(df.dtypes)

```

Poster_Link	object
Series_Title	object
Released_Year	object
Certificate	object
Runtime	object
Genre	object
IMDB_Rating	float64
Overview	object
Meta_score	float64
Director	object
Star1	object
Star2	object
Star3	object
Star4	object
No_of_Votes	int64
Gross	object
dtype:	object

Menghitung Hubungan Antar Data

Kami menganalisis hubungan antar variabel numerik untuk mengidentifikasi faktor-faktor yang mungkin berpengaruh signifikan dalam sistem penyaluran film.

```
correlation = df.corr(numeric_only=True)
print(correlation)
```

[10] ✓ 0.0s

	IMDB_Rating	Meta_score	No_of_Votes
IMDB_Rating	1.000000	0.268531	0.494979
Meta_score	0.268531	1.000000	-0.018507
No_of_Votes	0.494979	-0.018507	1.000000

Mendapatkan Ringkasan Statistik

Kami menganalisis ringkasan statistik dari data numerik dan non-numerik untuk memahami distribusi dan karakteristik data, yang penting untuk eksplorasi awal dalam pengembangan sistem penyaluran film.

```
df.describe()
```

[11] ✓ 0.0s

	IMDB_Rating	Meta_score	No_of_Votes
count	1000.000000	843.000000	1.000000e+03
mean	7.949300	77.971530	2.736929e+05
std	0.275491	12.376099	3.273727e+05
min	7.600000	28.000000	2.508800e+04
25%	7.700000	70.000000	5.552625e+04
50%	7.900000	79.000000	1.385485e+05
75%	8.100000	87.000000	3.741612e+05
max	9.300000	100.000000	2.343110e+06

Menganalisis Distribusi Kategori

Kami menghitung frekuensi setiap kategori film dan menganalisis distribusinya untuk mendapatkan gambaran tentang jenis film yang paling umum dan yang lebih jarang.

```
> ✓ df['Genre'].value_counts()
12] ✓ 0.0s

.. Genre
Drama 85
Drama, Romance 37
Comedy, Drama 35
Comedy, Drama, Romance 31
Action, Crime, Drama 30
..
Adventure, Drama, Musical 1
Drama, Fantasy, War 1
Comedy, Music 1
Drama, Music, Mystery 1
Comedy, Mystery, Romance 1
Name: count, Length: 202, dtype: int64
```

Data Preparation

Penyaringan Informasi

Dari sekumpulan data awal yang luas, kami memilih fokus pada informasi yang paling penting untuk sistem penyararan film kami. Kami memutuskan untuk menggunakan data berikut:

- Judul Seri
- Tahun Rilis
- Sertifikat
- Peringkat IMDB
- Skor Meta
- Jumlah Suara
- Pendapatan Kotor

Pemilihan ini didasarkan pada pentingnya data tersebut dalam membuat saran yang akurat, mencakup informasi utama seperti judul, tahun penayangan, penilaian, dan tingkat popularitas film. Dengan berfokus pada data ini, kami dapat mengembangkan sistem penyararan yang efektif dan efisien.

Penyempurnaan Data

Untuk mempertajam analisis, kami menyaring data dengan menghilangkan informasi yang kurang relevan seperti tautan poster, durasi, genre, sinopsis, sutradara, dan daftar pemain. Kami mempertahankan informasi kunci seperti judul film, tahun rilis, sertifikasi, penilaian

IMDB, skor Meta, jumlah suara, dan pendapatan kotor. Langkah ini membantu kami mengoptimalkan dataset untuk pengembangan sistem penyarhan yang lebih fokus dan efisien.

```
column_drop = ['Poster_Link', 'Runtime', 'Genre', 'Overview', 'Director', 'Star1', 'Star2', 'Star3', 'Star4']
df.drop(column_drop, inplace=True, axis=1)

df.head()
```

4] ✓ 0.0s

	Series_Title	Released_Year	Certificate	IMDB_Rating	Meta_score	No_of_Votes	Gross
0	The Shawshank Redemption	1994	A	9.3	80.0	2343110	28,341,469
1	The Godfather	1972	A	9.2	100.0	1620367	134,966,411
2	The Dark Knight	2008	UA	9.0	84.0	2303232	534,858,444
3	The Godfather: Part II	1974	A	9.0	90.0	1129952	57,300,000
4	12 Angry Men	1957	U	9.0	96.0	689845	4,360,000

+ Code + Markdown

Pengecekan Data Kosong

Kami memeriksa tipe data setiap kolom dan menghitung jumlah data yang hilang. Ini penting untuk memastikan kesesuaian tipe data dan mengidentifikasi kolom-kolom yang mungkin memerlukan penanganan khusus dalam tahap pra-pemrosesan.

```
df.isna().sum()
```

[24] ✓ 0.0s

```
Series_Title      0
Released_Year     0
Certificate       101
IMDB_Rating       0
Meta_score       157
No_of_Votes       0
Gross            169
dtype: int64
```

Penghapusan Data Tidak Lengkap

Dalam proses penyempurnaan data, kami menghitung jumlah baris awal, menghapus baris dengan nilai kosong, dan menghitung jumlah baris yang tersisa serta yang dihapus. Langkah ini penting untuk memastikan kualitas dan kelengkapan data yang akan digunakan dalam sistem penyarhan film, meningkatkan keandalan analisis dan hasil rekomendasi.

```
print(f"Jumlah baris sebelum penghapusan: {len(df)}")
df_cleaned = df.dropna()
print(f"Jumlah baris setelah penghapusan: {len(df_cleaned)}")
print(f"Jumlah baris yang dihapus: {len(df) - len(df_cleaned)}")

[5] ✓ 0.0s

... Jumlah baris sebelum penghapusan: 1000
    Jumlah baris setelah penghapusan: 714
    Jumlah baris yang dihapus: 286
```

Kesimpulan

Dalam kegiatan ini, kami telah menjalankan beberapa tahap krusial dalam pengelolaan dan penelaahan data untuk merancang sistem rekomendasi film. Diawali dengan mengunduh kumpulan data dari Kaggle, kami melakukan penjelajahan awal menggunakan Pandas guna memahami susunan dan ciri-ciri data. Proses ini meliputi pemeriksaan jenis data, analisis statistik ringkas, dan pengenalan nilai yang tidak ada.

Berikutnya, kami melaksanakan penyaringan data dengan menitikberatkan pada ciri-ciri yang paling berkaitan dengan rekomendasi film, seperti judul, tahun tayang, penilaian, dan jumlah pemilih. Proses pembersihan data, termasuk penghapusan baris yang tidak lengkap, menjamin mutu data yang akan dimanfaatkan dalam sistem.

Melalui rangkaian analisis ini, kami telah menyiapkan kumpulan data yang kokoh dan relevan untuk pengembangan sistem rekomendasi film yang efisien. Langkah-langkah ini meletakkan fondasi yang kuat untuk tahap selanjutnya dalam pembuatan algoritma rekomendasi yang tepat dan bermanfaat bagi pengguna.