

ANALISIS REKOMENDASI FILM DARI DATA IMDB MENGUNAKAN PYTHON

ANALYSIS OF FILM RECOMMENDATIONS FROM IMDB DATA USING PYTHON

Rahmi Putri Kurnia¹, Yori Adi Atma²

^{1,2})Politeknik Negeri Padang

*CorrespondingEmail:rahmiputrik@gmail.com

ABSTRAK

Penelitian ini memanfaatkan penggunaan beberapa library yang ada pada bahasa pemrograman Python dalam menentukan rekomendasi film dari situs IMDb sebagai sumber data dengan berdasarkan pada kategori atau genre film yang diinginkan / dicari oleh pengguna situs IMDb. Penentuan rekomendasi film ini sering kali menjadi kebingungan bagi pecinta film dalam menentukan pilihannya pada film apa yang akan dinikmati. Penelitian ini menggunakan bahasa pemrograman Python dengan librarynya digunakan untuk menganalisis dan menyelesaikan masalah, dimana ada terlalu banyak film yang tersedia di IMDb. sehingga akan menghasilkan output berupa rekomendasi film yang dapat menjadi acuan bagi pecinta film pengguna situs IMDb.

Kata Kunci : Rekomendasi, IMDB, Python, Library, Analisis.

ABSTRACT

This research utilizes the use of several existing libraries in the Python programming language in determining film recommendations from the IMDb site as a data source based on the category or genre of films desired/searched for by IMDb site users. The determination of film recommendations is often confusing for film lovers in making choices about which films to enjoy. This study uses the Python programming language with its library used to analyze and solve problems, where there are too many films available on IMDb. So that it will produce output in the form of film recommendations which can be a reference for film lovers using the IMDb site.

Keywords: Recommendations, IMDB, Python, Libraries, Analytics

PENDAHULUAN

Di tengah sibuknya kehidupan, hiburan menjadi salah satu item penting yang dapat menjadi penyeimbang kestabilan hidup. Salah satu hiburan yang bisa dijadikan pilihan yaitu film. Saat ini, dapat dilihat bahwa perkembangan industri film sangat berkembang pesat. Industri perfilman terus menerus tidak berhenti memproduksi film dengan berbagai jenis. Selain itu adanya faktor ketertarikan pecinta film yang berbeda satu dan lainnya. Hal-hal inilah yang tentunya terkadang membuat pecinta film menjadi bingung dalam menentukan tontonannya.

IMDb adalah salah satu dari situs yang telah cukup lama menyediakan informasi-informasi terkait hal diatas. Karena telah beroperasi cukup lama, sehingga memiliki jumlah data yang sudah cukup besar. Sehingga banyak pengguna atau user yang kebingungan dengan film apa yang akan di tonton.

Pada penelitian ini penulis akan melakukan analisa terkait dataset IMDB yang didapatkan dari Kaggle yang akan diolah dan digunakan sebagai acuan untuk membuat rekomendasi film dengan menggunakan fungsi-fungsi yang ada pada bahasa pemrograman python.

Python merupakan sebuah bahasa pemrograman yang dapat digunakan dalam proses klusterisasi dataset yang

dimiliki sebagai data mentah. Dengan hasil, sebagai bentuk rekomendasi atau grafik yang mudah dibaca dan dimengerti oleh orang awam.

Mendapatkan rekomendasi film yang diinginkan adalah tujuan pengguna mengakses situs IMDb. Dan untuk mendapatkan hasil rekomendasi yang terurut dan terdata dengan baik, penulis akan membuat analisis berdasarkan dataset IMDb yang didapatkan dari situs kaggle.

Berdasarkan uraian pendahuluan tersebut maka penelitian ini mengangkat tema dengan judul “Analisis Rekomendasi Film Imdb Menggunakan Python”.

METODE PENELITIAN

A. IMDb

IMDb adalah singkatan dari Internet Movie Database yang merupakan sebuah situs web yang menyediakan informasi mengenai film, acara televisi serta informasi terkait orang-orang yang terlibat.

IMDb pada awalnya berasal dari postingan penggemar sebuah film, kemudian orang dengan minat yang sama merespon dengan menambahkan daftar yang berbeda, sehingga akhirnya daftar itu pun terus berkembang dan IMDb dirilis pada tahun 1990 dan diakuisisi oleh amazon di tahun 1998. Hingga saat ini situs IMDb terus berkembang.

B. Python

Python awalnya dibuat pada akhir tahun 1980 oleh Guido van Rossum, Belanda. Implementasi bahasa pemrograman ini dimulai pada akhir tahun 1989. Pada tahun 2000, python 2.0 dirilis dengan banyak fitur utama yang baru. karena sejumlah besar kode tidak dapat dengan mudah dilanjutkan ke Python 3, maka tanggal akhir pemakain versi ini yang sebelumnya hanya hingga 2015 diperpanjang menjadi hingga tahun 2020.

Pada 2022, perilsan Python 3.10.4 dan 3.9.12 dipercepatdikarenakan banyak masalah keamanan. Dan kini Python menjadi salah satu bahasa pemrograman yang cukup banyak dipelajari dan digunakan.

Pada bagian ini, penulis menjelaskan tentang metodologi penelitian yang digunakan dalam menganalisis dan menyelesaikan masalah penelitian yang dihadapi. Metodologi penelitian ini sangat membantu penulis dalam proses penyelesaian masalah. Penelitian ini memiliki beberapa tahapan dalam pelaksanaan kegiatan yang tertuang pada kerangka kerja penelitian yaitu identifikasi masalah, analisis masalah, menentukan tujuan, mempelajari literatur, mengumpulkan data, menganalisis data dengan memanfaatkan library yang ada

pada python.

Dalam metodologi penelitian ada urutan kerangka kerja yang harus diikuti, urutan kerangka kerja ini merupakan langkah-langkah yang dilakukan dalam penulisan. Adapun uraian langkah kerjanya adalah sebagai berikut :

a. Mengidentifikasi Masalah

Dalam tahap ini, penulis melakukan peninjauan ke sistem yang akan diteliti, dalam hal ini adalah IMDb. Guna untuk mengamati serta melakukan eksplorasi dan menggali permasalahan yang ada. Proses ini menjadi tahap awal untuk menentukan rumusan masalah dari penelitian.

b. Menganalisis Masalah

Tahap ini, penulis mencoba memahami masalah - masalah yang muncul setelah ditentukan ruang lingkup atau batasannya. Hal ini berguna agar penelitian ini berjalan fokus pada masalah - masalah tertentu saja. Dengan melakukan penganalisaan pada masalah yang telah ditentukan tersebut, diharapkan masalah dapat dipahami dengan baik dan dapat teranalisa dengan sebaik mungkin.

c. Menentukan Tujuan

Berdasarkan dari proses sebelumnya mengenai pemahaman masalah yang dianalisa, maka langkah berikutnya adalah menentukan tujuan yang akan

dicapai dari penelitian ini. Pada tujuan ini ditentukan target yang akan dicapai, terutama yang dapat mengatasi masalah-masalah yang ada.

d. Mempelajari Literatur

Literatur-literatur yang dipakai sebagai bahan referensi dalam penelitian ini adalah dari jurnal-jurnal ilmiah khususnya jurnal dan artikel ilmiah yang membahas tentang rekomendasi film dan penggunaan python. Literatur-literatur ini akan menjadi pedoman untuk melakukan penelitian agar memudahkan proses penelitian.

e. Mengimplementasikan Library Python

Pada tahap ini penulis melakukan implementasi penggunaan library python untuk mengolah data yang telah didapatkan.

f. Menganalisis Hasil Pengujian

Pada tahap ini akan dilakukan pengamatan dan analisis terhadap hasil pengujian dalam menentukan hasil rekomendasi film berdasarkan data yang ada.

mengambil dataset IMDb dari situs Kaggle yang dapat dipercaya.

Sebelum data diolah terlebih dahulu ditentukan kategori pengeneretan untuk hasil rekomendasi. Selanjutnya penulis menentukan library-library yang akan digunakan dalam mengolah data.

Untuk memudahkan pembacaan data, data dapat dianalisis untuk menerangi bagian-bagian penting atau untuk mencari sebuah pola. Analisis data dapat dilakukan dengan mencari komponen-komponen statistik penting dari data. Komponen yang akan kita analisis dari data ini berupa jumlah data, rata-rata(mean), range, persentil, serta nilai ekstremum yaitu nilai maksimum dan minimum. Jumlah data menunjukkan jumlah data yang terdapat pada dataset.

2. Tahapan Pengolahan Data

Pada tahap ini penulis melakukan pengolah dataset IMDb dengan menggunakan library-library python, seperti : pandas, numpy dan matplotlib.

Program library Pandas sendiri dapat dijalankan menggunakan berbagai text editor, namun sangat disarankan untuk menggunakan Jupyter Notebook. Hal ini disebabkan karena kemampuan Jupyter untuk menjalankan kode dari satu cell tertentu tanpa perlu menjalankan file secara keseluruhan. Selain itu, Jupyter juga menyediakan cara yang mudah untuk

HASIL DAN PEMBAHASAN

1. Analisa Data

Untuk memudahkan proses ini, penulis membutuhkan data yang real yaitu data yang dapat dipertanggungjawabkan kebenarannya, untuk itu penulis

memvisualisasikan *data frames* dan plot yang dibuat oleh Pandas.

3. Implementasi dan Pengujian

Pada tahap ini penulis akan menjabarkan penggunaan library python dalam pengolah dataset.

Langkah-langkah yang dilakukan :

- a. Mengimpor dataset yang akan digunakan.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from google.colab import files
filenya = files.upload()
```

Choose Files No file chosen Upload widget is on Saving movie_metadata.csv to movie_metadata.csv

proses import selesai jika sudah menunjukkan alert data berhasil diimport. Hal ini berarti kita sudah dapat memanggil data tersebut seperti gambar berikut :

```
movies = pd.read_csv('movie_metadata.csv')
movies.head(3)
```

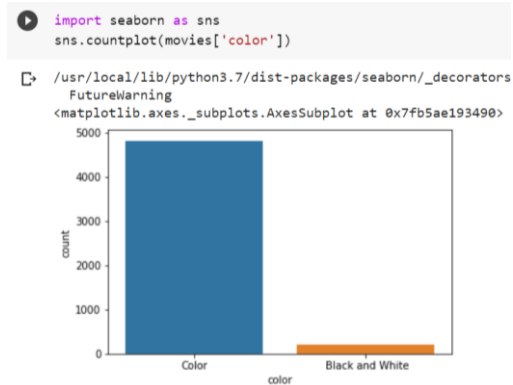
	color	director_name	num_critics_for_reviews	duration	director_facebook_likes	actor_3_facebook_likes	actor_2_name
0	Color	James Cameron	723.0	178.0	0.0	855.0	Joe David Moore
1	Color	Gore Verbinski	302.0	169.0	563.0	1006.0	Orlando Bloom
2	Color	Sam Mendes	602.0	148.0	0.0	181.0	Rory Kinnear

Untuk mendapatkan data - data yang diinginkan kita bis membuat perintah khusus, agar data yang ditampilkan sesuai dengan yang kita inginkan. Contohnya dapat dilihat pada gambar berikut :

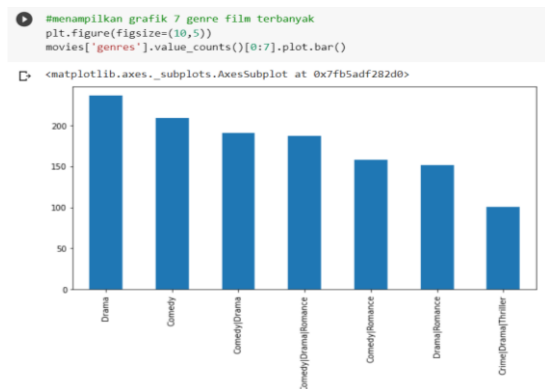
```
#mengonversi file berdasarkan skor
movies[movies.sort_values('imdb_score', ascending=False)
#tampilkan 20 barisan
movies[['movie_title', 'title_year', 'director_name', 'genres', 'imdb_score']].head(20)
```

	movie_title	title_year	director_name	genres	imdb_score
2795	Towering Inferno	NaN	John Blanchard	Comedy	9.5
1937	The Shawshank Redemption	1994.0	Frank Darabont	CrimeDrama	9.3
3489	The Godfather	1972.0	Francis Ford Coppola	CrimeDrama	9.2
4409	Koolhaas: Vengance	2016.0	John Stockwell	Action	9.1
2624	Dakota	NaN	NaN	Drama	9.1
3397	Dakota	NaN	NaN	Drama	9.1
66	The Dark Knight	2008.0	Christopher Nolan	ActionCrimeDramaThriller	9.0
2837	The Godfather Part II	1974.0	Francis Ford Coppola	CrimeDrama	9.0
3481	Fargo	NaN	NaN	CrimeDramaThriller	9.0
239	The Lord of the Rings: The Return of the King	2003.0	Peter Jackson	ActionAdventureDramaFantasy	8.9
4822	12 Angry Men	1957.0	Sidney Lumet	CrimeDrama	8.9

- b. Menggunakan berbagai library sesuai kebutuhan untuk melakukan analisa dan pengolahan data.



Gambar 3. Hasil penggunaan library seaborn



Gambar 4. Hasil Plot Data

Dengan penggunaan library seperti diatas akan lebih memudahkan pengguna situs dalam melihat hasil rekomendasi film. Rekomendasi film pun terlihat lebih terstruktur dan mudah dimengerti.

Hasil dari penggunaan library python dapat dilihat dari hasil yang didapatkan. Dari data mentah yang sangat sulit dibaca dapat diolah menjadi sebuah informasi. Dan dengan menggunakan library yang ada pada python, hasil pembacaan data menjadi lebih mudah di baca oleh orang awam.

SIMPULAN

Berdasarkan analisa data dan pengujian dengan menggunakan library pada python, maka dapat diambil kesimpulan sebagai berikut :

- a. Pada penelitian ini penulis menggunakan beberapa library dalam penyajian data untuk memudahkan analisa dan mendapatkan hasil rekomendasi film dari dataset.
- b. Masing-masing library memiliki fungsi yang berbeda, sehingga penentuan penggunaan library sangat menentukan pada hasil yang ditampilkan.
- c. Dengan menggunakan library ada python dapat memberikan informasi yang lebih mudah dibaca dan dimengerti oleh orang awam.

DAFTAR PUSTAKA

- A. Ambarwati, Q. J. Adrian and Y. Herdiyeni, 2020. ***Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman***, Jurnal RESTI, vol. 4, no. 1, pp. 117-122.
- Bernadus Gunawan Sudarsono, Marcell Ignatius Leo, Ali santoso3) dan Felix Hendrawan (2021). ***Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner***. Journal of Business and Audit Information Systems VVol 4 (No.1).
- I. D. A P, 2016, ***Pengaruh Rating Dalam Situs IMDb Terhadap Keputusan Menonton Di Kota Bandung***, Bandung: Telkom University.
- Maylida Izattul Wardah, Syahrizal Dwi Putra (2022). ***Implementasi Machine Learning Untuk Rekomendasi Film Di Imdb Menggunakan Collaborative Filtering Berdasarkan Analisa Sentimen IMDB***. Jurnal Manajemen Informatika Jayakarta. Vol.2, No. 3.
- R. A. Abarja and A. Wibowo, 2020. ***Movie Rating Prediction using Convolutional Neural Network based on Historical Values***, International Journal of Emerging Trends in Engineering Research, vol. 8, no. 5, pp. 2156-2164.
- Sarosa Castrena Abadi, Muhammad Ayat Hidayat, Purwandito Tulus Asmoro (2020). ***Sistem Rekomendasi Film Berbasis Jejaring Sosial (Twitter) Menggunakan Ibm Bluemix***. Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM). Vol. V, No. 1.
- V. R. Prasetyo, B. Hartanto and A. A. Mulyono, 2019. ***Penentuan Pembimbing Tugas Akhir Mahasiswa Jurusan Teknik Informatika Universitas Surabaya Dengan Metode Dice Coefficient***, TEKNIKA, vol. 8, no. 1, pp. 44-51.