

Sistem pencarian judul film berbasis teks



UNIVERSITAS DIAN NUSWANTORO
SEMARANG

Nama : Mayra Anggraini

NIM : A11.2022.14765

Deskripsi Singkat

Proyek ini bertujuan untuk mengembangkan sebuah sistem mesin pencari yang dapat membantu pengguna menemukan judul film berdasarkan input teks, seperti kata kunci, genre, atau frasa tertentu yang diingat dari film tersebut. Sistem ini dirancang untuk mendukung pengguna yang mungkin lupa judul lengkap film atau ingin mencari film berdasarkan karakteristik spesifik. Mesin pencari ini dapat diimplementasikan sebagai bagian dari aplikasi atau situs web untuk pengguna umum.

Masalah dan tujuan yang ingin diselesaikan

Masalah utama yang ingin diselesaikan adalah bagaimana menyediakan akses yang cepat dan akurat bagi pengguna yang ingin menemukan judul film tertentu berdasarkan deskripsi atau kata kunci. Tujuan utama dari proyek ini adalah mengembangkan mesin pencari yang dapat segera digunakan untuk deployment dan terintegrasi dengan antarmuka pengguna (UI) yang mudah diakses. Dengan adanya mesin pencari ini, pengguna dapat mengakses database film dan menemukan judul yang mereka cari tanpa perlu browsing manual atau mencarinya secara acak.



Penjelasan datasets

Dataset yang digunakan berasal dari koleksi data film publik, seperti IMDb atau The Movie Database (TMDb). Dataset ini dapat diperoleh melalui API atau scraping data dari situs yang relevan. Setiap entri dalam dataset akan menyimpan berbagai fitur penting, seperti:

- **Judul Film:** Nama film yang dapat digunakan sebagai kata kunci utama.
- **Genre:** Kategori film (misalnya: drama, komedi, aksi).
- **Deskripsi Singkat:** Sinopsis pendek atau frasa yang relevan dengan film.
- **Pemeran Utama:** Nama aktor atau aktris yang berperan penting dalam film.
- **Tanggal Rilis:** Tahun dan bulan film tersebut dirilis.

Alur

Input Data dan Persiapan:

-  Dataset film dikumpulkan dan disimpan dalam bentuk .csv atau JSON, kemudian diolah untuk menghilangkan noise atau data duplikat.
-  Fitur utama seperti judul, genre, deskripsi, dan pemeran diekstraksi untuk digunakan sebagai dasar pencarian.

Feature Engineering:

- • Ekstraksi fitur teks dari kolom deskripsi, judul, dan genre dilakukan agar setiap film dapat diidentifikasi berdasarkan karakteristik spesifiknya.

Proses Pencarian:

- ☐ Saat pengguna memasukkan kata kunci, sistem akan melakukan pencarian menggunakan teknik pencocokan teks untuk menemukan film yang paling relevan.
- ☐ Algoritma seperti cosine similarity dapat digunakan untuk menghitung kemiripan antara kata kunci input dan data deskripsi film.

Output:

- • Sistem akan menampilkan daftar film yang memiliki tingkat relevansi tinggi dengan input pengguna.
- Output dapat berupa daftar film yang menampilkan judul, tahun rilis, dan genre film.