

Nama : STEVEN KURNIAWAN , VIOLA GLINKO

NIM : 211110607 , 211111906

JUDUL : REKOMENDASI FILM BERDASARKAN RATING

1) PEMBAGIAN TUGAS :

STEVEN KURNIAWAN : Training Data dan Pembuatan Model KNN dan SVM , Penjelasan keseluruhan proses prepare dataset hingga running model

VIOLA GLINKO : Pengumpulan dan pengolahan dataset , seperti preparing , cleaning , picking dan merge dataset hingga menjadi sebuah dataset final yang utuh , Penjelasan proses deployment dan perancangan kerangka website html , app.py

2. DESKRIPSI MASALAH YANG DIANGKAT :

Minat Masyarakat dalam menonton film semakin kurang karena tidak adanya system yang mengklasifikan Tingkat rekomendasi film berdasarkan genre favorit setiap individu.

3. MODEL YANG DITERAPKAN :

Model K-Nearest Neighbour dan Support Vector Machine :

KNN (K-Nearest Neighbors): KNN sangat cocok untuk rekomendasi film karena kemampuannya untuk menemukan "tetangga" terdekat, atau film-film yang mirip, berdasarkan kesamaan fitur seperti genre, durasi, sutradara, dan rating. Dalam KNN, setiap film baru yang diinput akan dibandingkan dengan data film lainnya, dan film-film dengan kesamaan tertinggi (misalnya, berdasarkan preferensi pengguna atau karakteristik film) akan direkomendasikan. Ini membuat KNN sangat intuitif untuk rekomendasi berbasis konten, di mana film yang mirip dengan pilihan pengguna sebelumnya disarankan tanpa memerlukan model pembelajaran yang kompleks.

SVM (Support Vector Machine): SVM juga bisa diterapkan untuk rekomendasi. SVM bekerja dengan menemukan hyperplane yang dapat memisahkan data ke dalam beberapa kategori berdasarkan fitur film atau preferensi pengguna. Dengan begitu, SVM cocok untuk rekomendasi film berbasis kategori di mana model dilatih untuk memisahkan film yang disukai dari yang tidak disukai, atau memprediksi genre tertentu yang cenderung disukai pengguna. SVM cenderung efektif dalam dataset rekomendasi yang memiliki fitur kompleks dan dapat digunakan untuk membangun rekomendasi film berbasis personalisasi yang lebih akurat.

4. input, proses, dan output yang dihasilkan

Customer memilih genre maksimal tiga , dengan durasi film hanya 1 – 3 jam , kemudian system akan otomatis men-generate prediksi score untuk menentukan rekomendasi film yang telah diolah oleh model dan dataset yang sudah di training

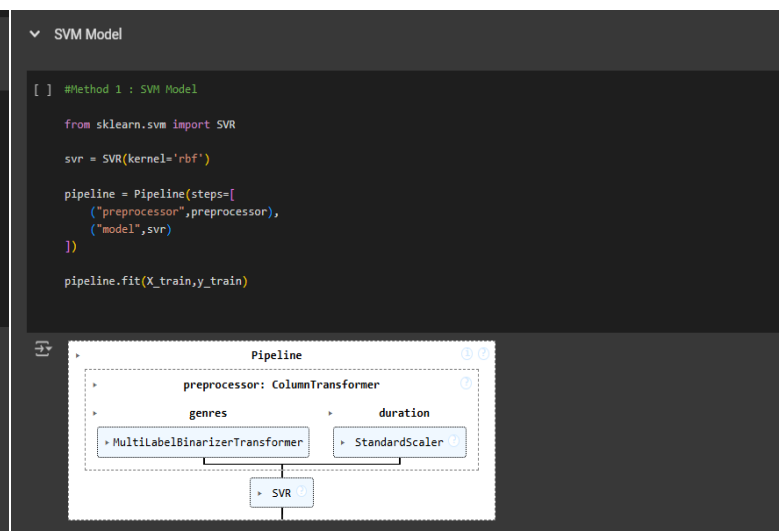
5. Kompleksitas Masalah dan Jumlah dataset

Kompleksitas masalah terdapat pada banyaknya source dataset yang kemudian harus di merge menjadi sebuah final dataset. Dataset diambil dari : Amazon_prime_title.csv , Disney_plus_title.csv , netlix_title.csv , imdb_movie.csv , imdb_top_1000.csv

6. Cara Preprocessing data dan Model yang Digunakan

proses preprocessing data dilakukan melalui pipeline menggunakan objek preprocessor, yang dirangkai dengan model KNN atau SVM. Berikut adalah cara singkat pra-pemrosesan yang diterapkan:

1. **Pra-pemrosesan dalam Pipeline:** Pipeline memungkinkan otomatisasi proses pra-pemrosesan sehingga data siap untuk pelatihan model. Dengan pipeline, data X_{train} dan X_{test} diproses menggunakan preprocessor terlebih dahulu sebelum dilatih pada model atau digunakan untuk prediksi.
2. **Langkah-Langkah Pra-Pemrosesan Umum:**
 - **Encoding Kategori:** Variabel kategori dalam data diubah menjadi format numerik (misalnya, menggunakan *OneHotEncoder* atau *LabelEncoder*).
 - **Scaling (Penskalakan Fitur):** Fitur numerik mungkin distandarkan atau dinormalisasikan menggunakan *StandardScaler* atau *MinMaxScaler* agar berada dalam rentang nilai tertentu. Scaling ini penting untuk model seperti SVM dan KNN yang sensitif terhadap skala fitur.
 - **Imputasi Nilai Kosong:** Nilai yang hilang dalam data dapat diisi (imputasi) dengan rata-rata, median, atau metode lain sesuai konfigurasi preprocessor.
3. **Output:** Pipeline pra-pemrosesan memastikan bahwa data berada dalam bentuk numerik yang terstandarisasi, memungkinkan model KNN atau SVM untuk mengidentifikasi pola kesamaan dengan lebih akurat.



6. Akurasi dan Tingkat Error

A. KNN

Akurasi : [69.67272727]

Error :

- Mean Absolute Error : 5.440122487802349
- Root Mean Squared Error 6.924154308903426
- Mean Squared Error 47.94391289350588
- R2 Score : 0.17641203319878695

B. SVM

Akurasi : [67.36301302]

Error :

- Mean Absolute Error : 4.776165933122121
- Root Mean Squared Error 6.0700697018747425
- Mean Squared Error 36.84574618561773
- R2 Score : 0.36705806107876415

7. Teknologi yang digunakan

Web Based Program dan Model KNN dan SVM yang

