**Sistem Rekomendasi Film (content-based filtering)**

Content-based filtering adalah algoritma sistem rekomendasi yang memanfaatkan kesamaan dari data lain(misal: kesamaan genre, kata kunci, dan lainnya), percobaan kali ini membuat sistem rekomendasi film memanfaatkan library cosine similatiry dari Scikit-Learn, dataset diambil dari: <https://www.kaggle.com/datasets/tmdb/tmdb-movie-metadata?select=tmdb_5000_movies.csv>

Tutorial: <https://www.youtube.com/watch?v=1xtrIEwY_zY&t=6351s>

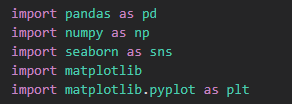
Github: <https://github.com/GilangAgungS/rekomendasi-film-content-based-filtering>

Untuk tampilan website kita membutuhkan API dari TMDB untuk menampilkan gambar dan judul dari hasil rekomendasi film.

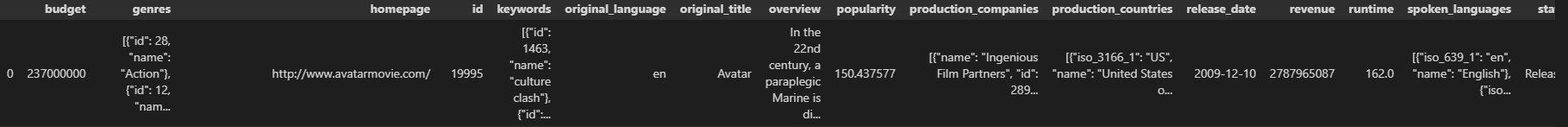
Tutorial API TMDB: <https://www.youtube.com/watch?v=mO3gvkiLkio>

Berikut adalah langkah-langkahnya:

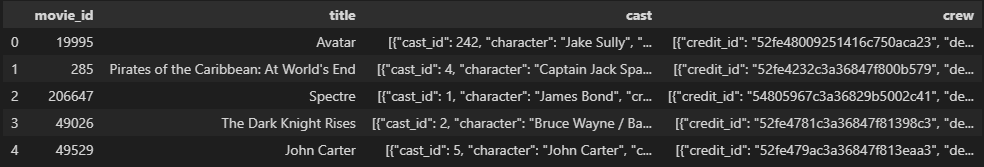
**1. Import library untuk mengolah dataset**



**2. Load dataset**





**3. Menggabungkan 2 dataset yang sudah di-load**



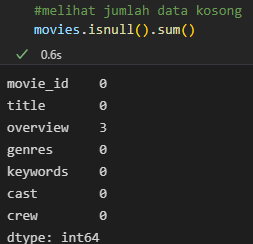
**4. Mengambil kolom yang diperlukan**

Kita ambil beberapa kolom yang dapat digunakan sebagai identitas film seperti movie\_id dan title, kolom berisi kata-kata yang nantinya dapat digabung sebagai kata kunci seperti kolom overview, genres, keywords, cast, dan crew.

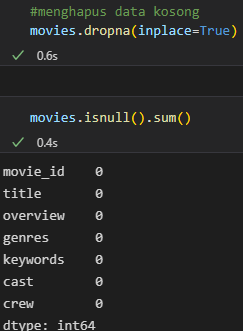


**5. Eksplorasi data**

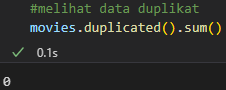
Melihat data yang kosong



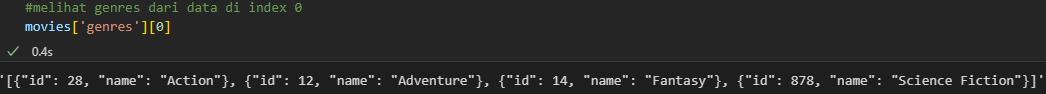
Terlihat pada kolom overview ada 3 data kosong, kita hapus data tersebut



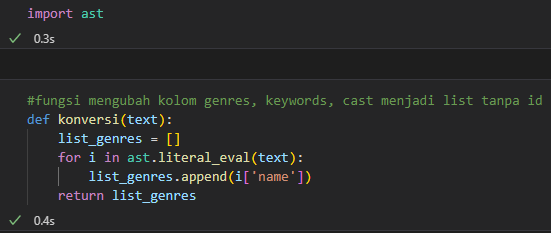
Cek data yang sama(duplikat)



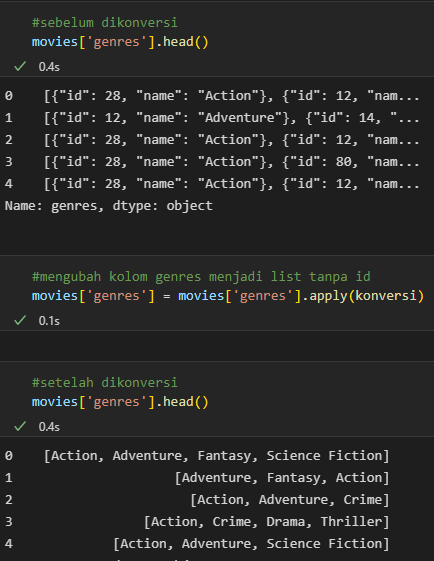
Berikut adalah contoh isi dari dari kolom genres, memiliki artibut id dan name dengan nilainya, kolom keyword, cast, dan crew memiliki bentuk data yang mirip, kita akan manipulasi data tersebut dan mengambil isi yang diperlukan.



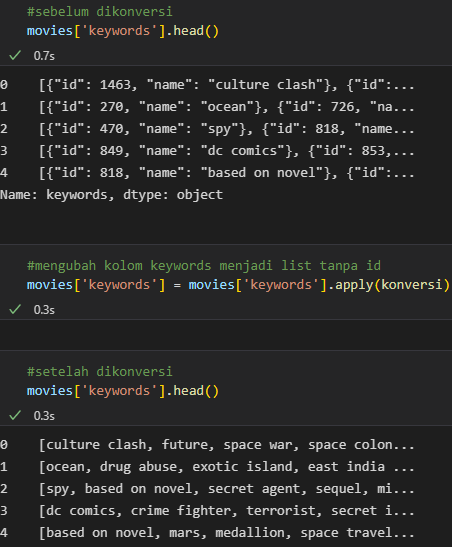
Buat fungsi untuk menghilangkan id dari data kolom genres, keyword, dan cast



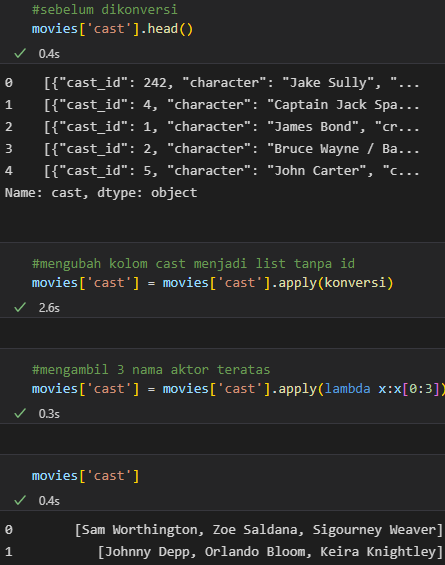
Konversi kolom genres



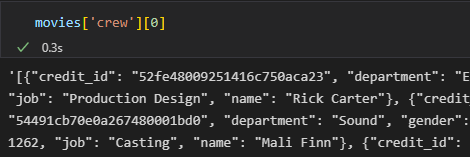
Konversi kolom keywords



Konversi kolom cast

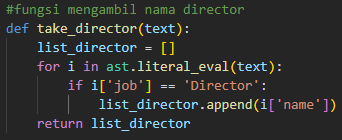


Melihat isi kolom crew

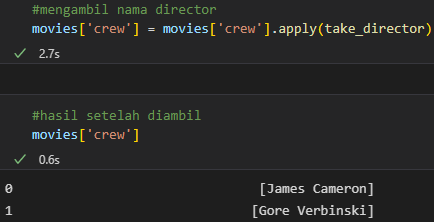


Kolom crew memiliki banyak sekali data, namun kita akan ambil nilai dari atribut director saja

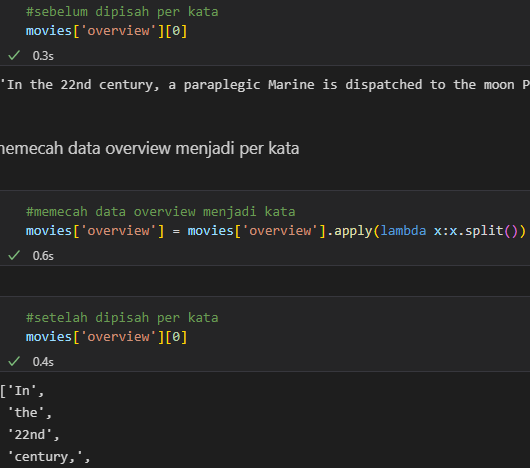
Fungsi untuk mengambil nilai dari atribut director



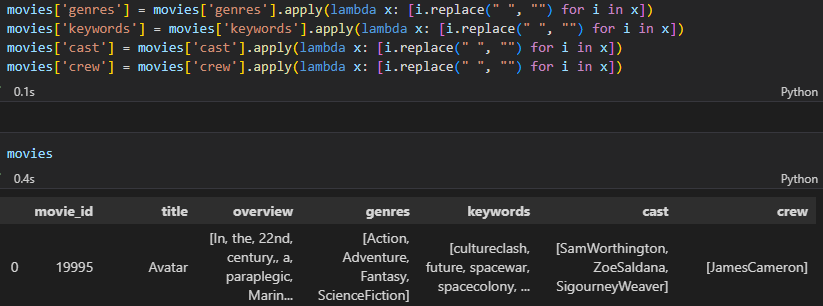
Hasil setelah diproses



Memecah isi kolom overview menjadi per kata yang dikumpulkan dalam list



Menghapus spasi dari kolom genres, keywords, cast, crew



Setelah modifikasi isi dari kolom genres, keywords, cast, dan crew, kita gabungkan seluruh isi kolom tersebut yang nantinya akan menjadi acuan dalam rekomendasi

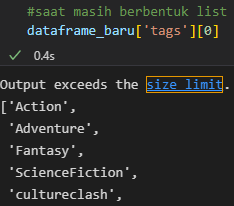
Menggabungkan isi kolom di atas menjadi satu dalam kolom baru bernama “tags”



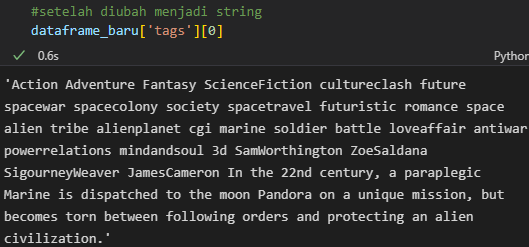
Kolom genres, keyword, cast, crew, dan overview sudah tidak digunakan karena isi kolom tersebut sudah digabung ke kolom tags, setelah itu kita buat variabel baru berisi kolom movie\_id, title, dan tags.



Mengubah kolom tags yang berisi list menjadi string utuh

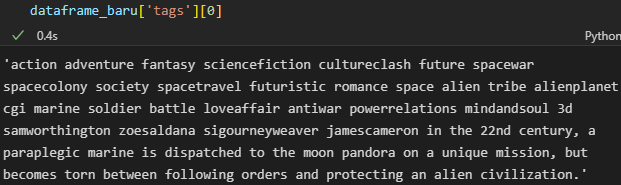






Mengubah isi kolom tags tanpa uppercase





**6. Membuat Vektor Rekomendasi**

CountVectorizer berfungsi untuk menghitung frekuensi kemunculan kata dan diubah menjadi vektor.

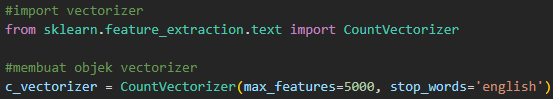
Sumber: <https://digilib.esaunggul.ac.id/public/UEU-Journal-20004-11_1244.pdf>

Max\_features berfungsi untuk membuat kolom berisi kata-kata, sehingga lebih cepat dalam menghitung jumlah kata yang muncul.

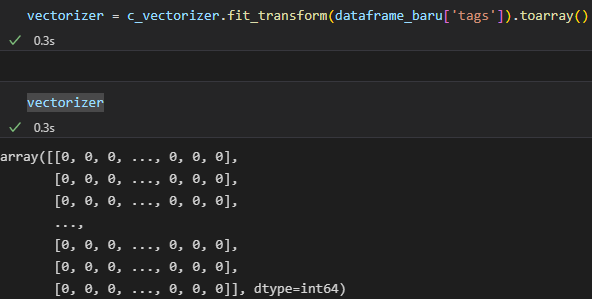
Sumber <https://www.megabagus.id/machine-learning-natural-language-processing/3/>

Stop\_words berfungsi untuk menghilangkan kata yang sering digunakan namun tidak bermakna seperti a, is, are dan sebagainya.

Sumber: <https://medium.com/sesuapnasi/memahami-fit-transform-pada-module-countvectorizer-d78bc8b8036f>



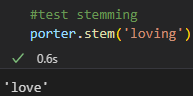
Proses vectorizer

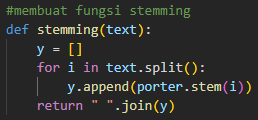


**7. Proses Stemming**

Stemming merupakan proses untuk mengubah kata-kata menjadi kata dasar

Sumber: <https://codepolitan.com/blog/stemming-word-dalam-carik-bot-59a9ef6e96088>


**8. Proses Cosine Similarity**

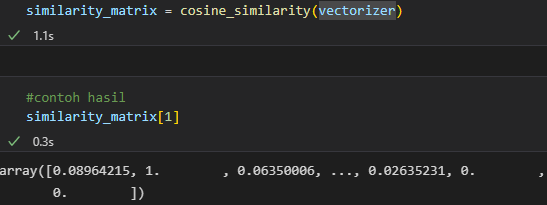
Cosine Similarity merupakan fitur dari Scikit-Learn untuk membandingkan kemiripan dari dokumen, dalam kasus ini adalah kesamaan kata dari kolom tags.

Sumber: <https://garudacyber.co.id/artikel/2205-pengertian-dan-penerapan-cosine-similarity>

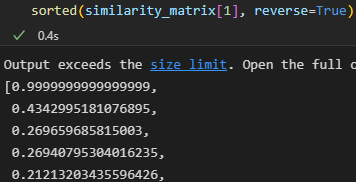


Membuat variabel “similarity\_matrix” berisi hasil perhitungan cosine\_similarity dari variabel vectorizer

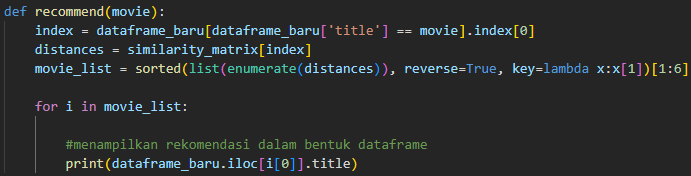




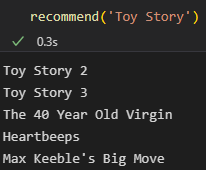
Contoh hasil similarity\_matrix diurutkan dari nilai terbesar



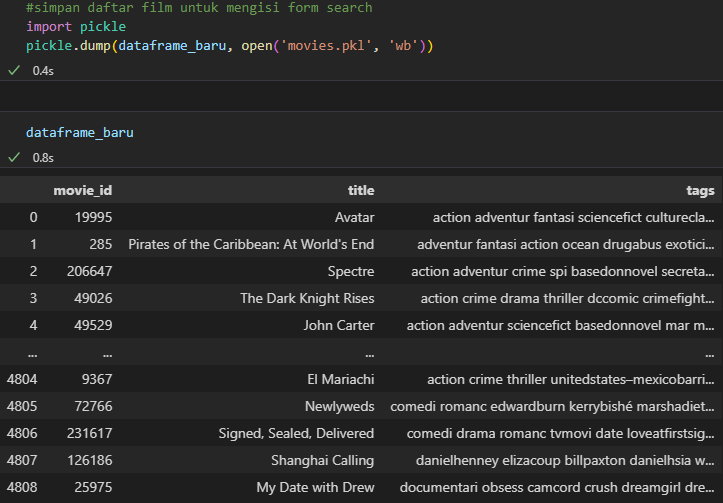
Membuat fungsi untuk menampilkan rekomendasi



Hasil rekomendasi



**9. Simpan Daftar Film dan Model Cosine Similarity**



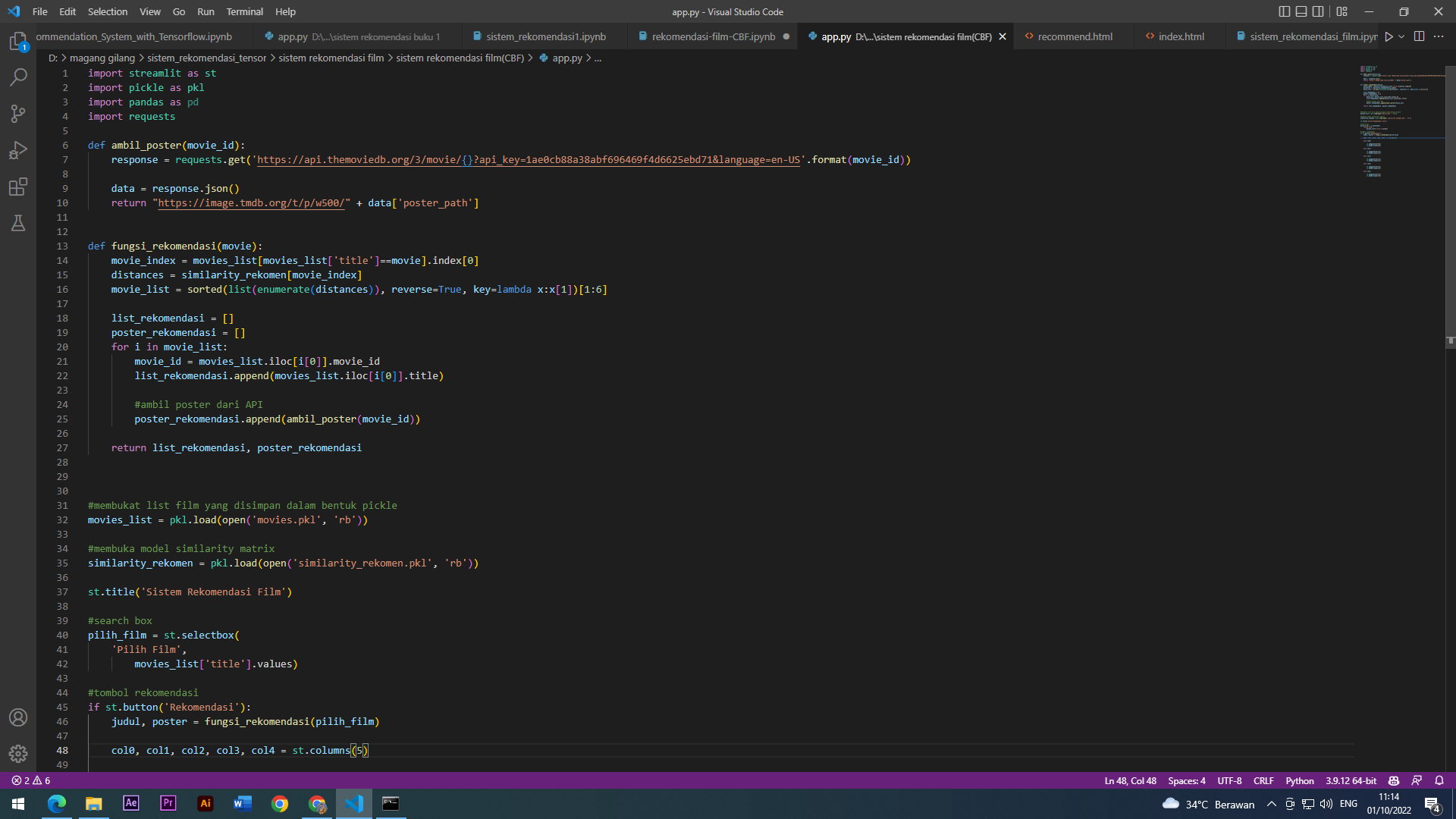
Simpan model

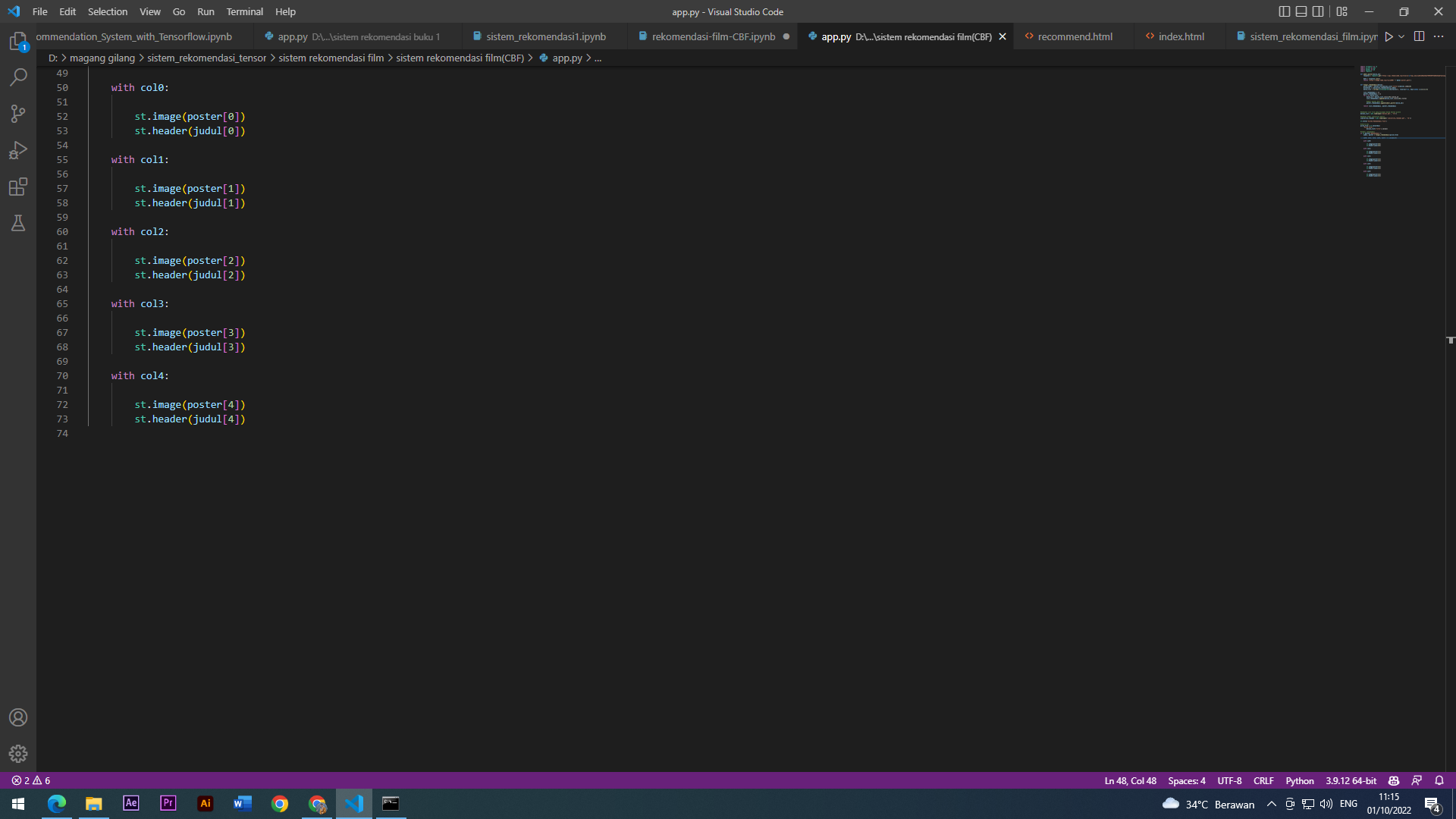


**10. Membuat Tampilan Website**

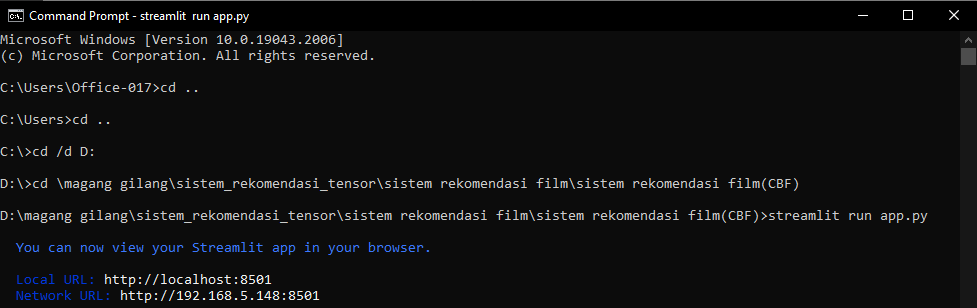
Pembuatan tampilan website menggunakan library Streamlit yang berjalan di bahasa python dan mempercepat pembuatan website. Sumber: <https://streamlit.io/>

Buat file dengan nama app.py, dan tulis kode seperti di bawah ini, modifikasi juga boleh

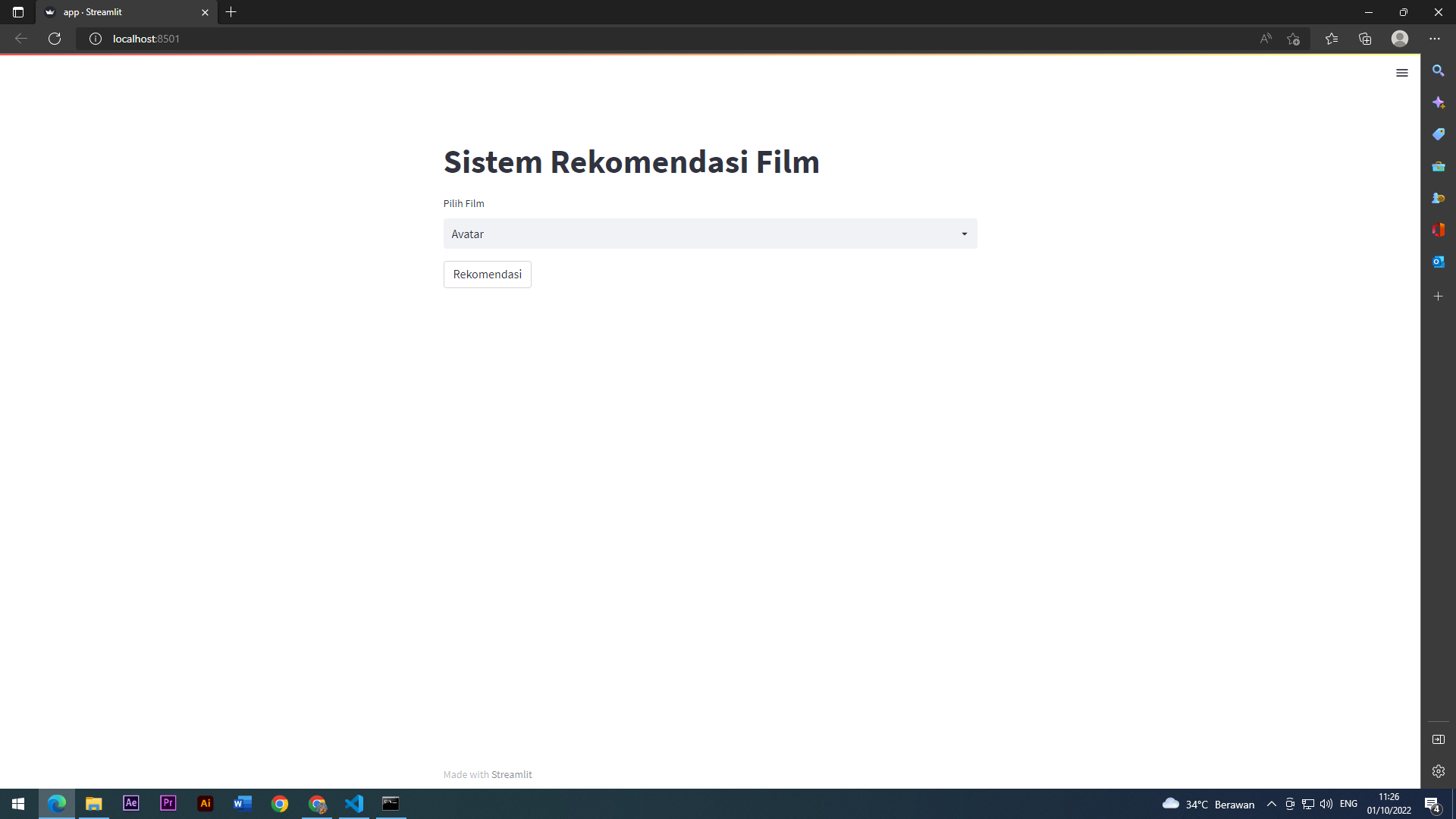




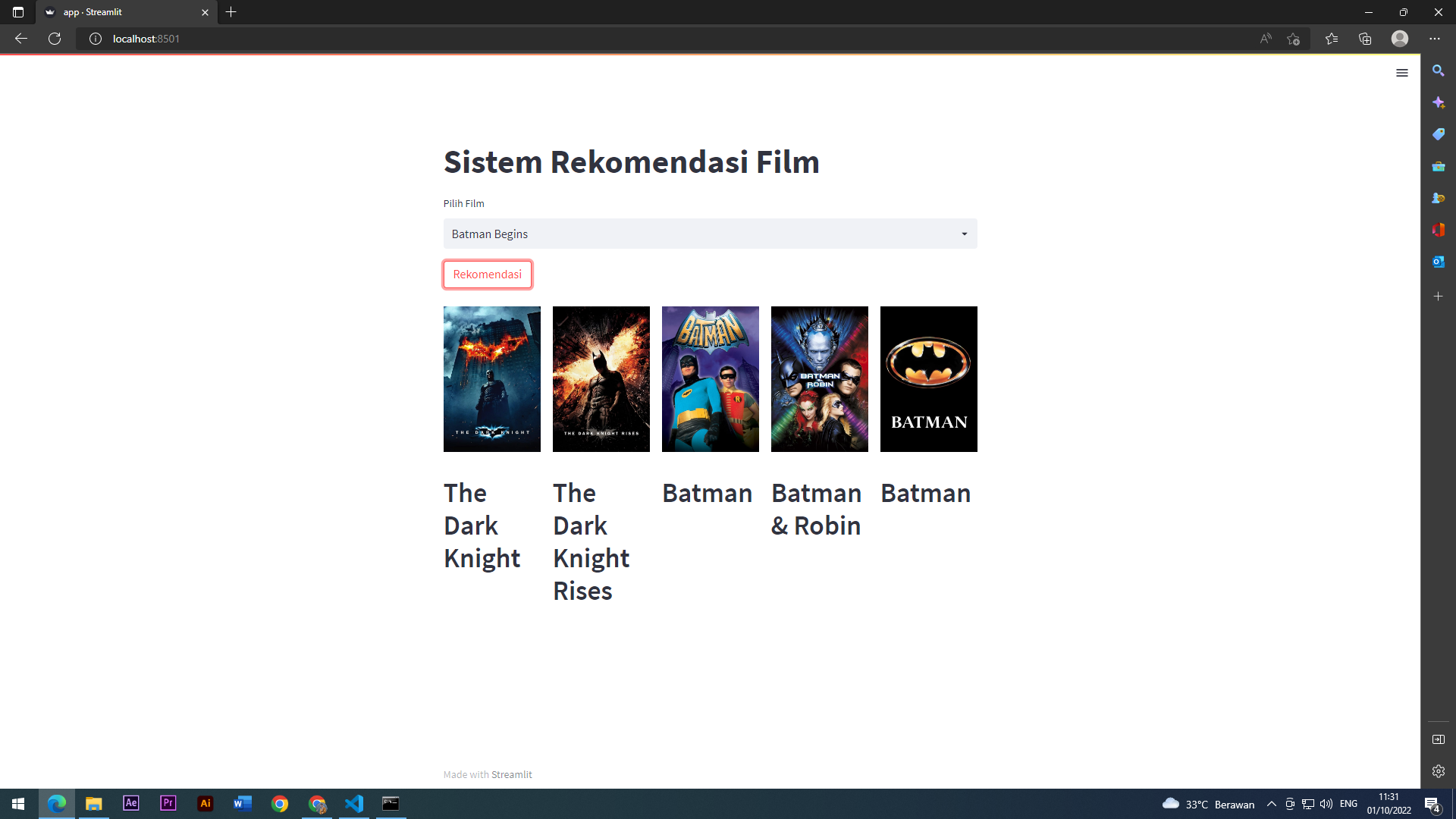
Buka Command Prompt, arahkan ke path tempat menyimpan app.py, lalu ketik “streamlit run app.py”



Hasilnya:



Contoh hasil rekomendasi film:

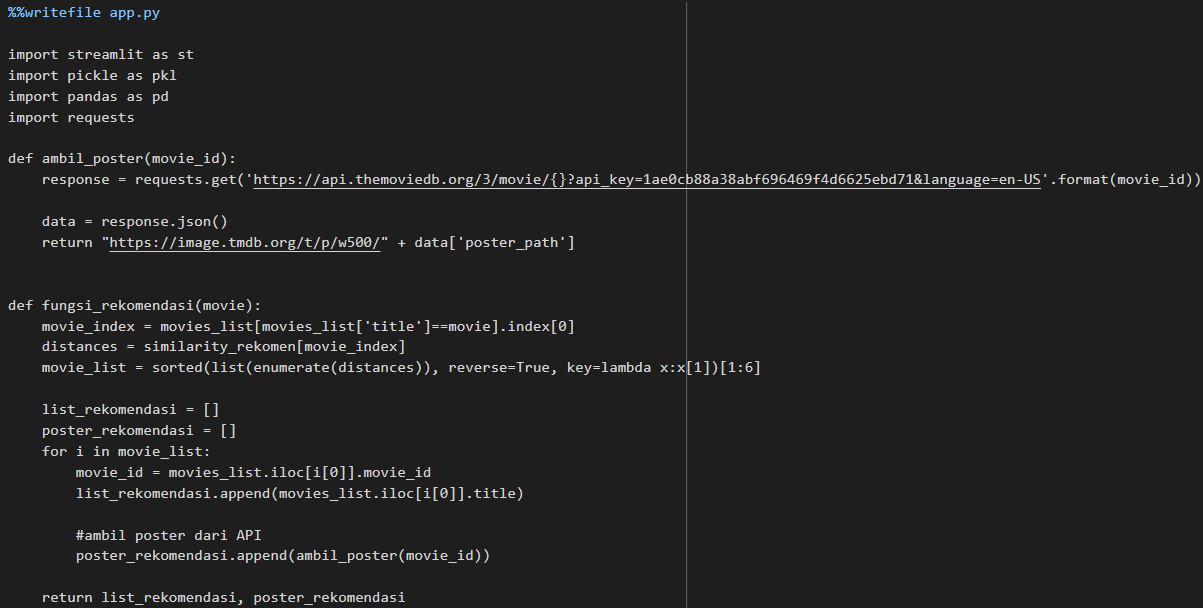


**11. Deploy dengan Mudah melalui Google Colab dan LocalTunnel**

*Install library* streamlit ke Google Colab

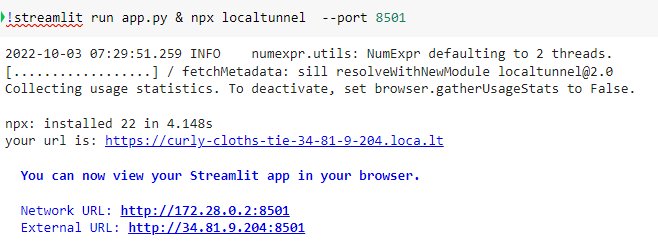


Upload file .ipynb dan file dataset ke Google Colab, lalu buat file app.py dan isi dengan kode “%%writefile app.py” dan lanjutkan dengan kode dari app.py di komputer local



\*kode terpotong

Jalankan seluruh cell di Google Colab, pastikan dapat berjalan dengan baik, kemudian ketik dan jalankan kode di bawah ini:



Tunggu hingga cell berjalan dan klik link yang muncul

Berikut adalah hasilnya:

