Movie Recommendation System

Изготвен от: Станимир, Стоил и Иван

Movie Recommendation Системата, която направихме работи на базата на над 4800 филма като се съпоставят според жанр, ключови думи, актьори и продуцент. Готовият резултат излиза под формата на графичен интерфейс като се показват 10 подобни на предварително указания от нас филм.

Избрахме да работим по Movie Recommendation System първо и най-важно защото и тримата бяхме съгласни на това. Имаше и други предложения, които пораждаха неволи сред нас и това беше решението по средата. Също така проекта, който сме изготвили не засяга важни за здравето теми, които бяха част от другите предложения, тъй като смятаме, че това е прекалено сериозна тема за работа и има заведения, в които човек да се погрижи за здравето си. Да обобщим, системата е с цел развлечение, лесна е за употреба и спестява от времето ни за търсене.

В проекта са използвани следните библиотеки:

numpy

Numpy или Numerical Python е библиотека за работа с масиви. Разликата между numpy и обикновените масиви в Python са, че тези на Numpy са по-лесно достъпни и позволяват по-бърза работа с тях(около 50 пъти по-бързи).

Картина, която съдържа текст, монтиран

Описанието е генерирано автоматично

scikit-learn

Една от най-важните библиотеки при работата с machine learning. Тя се използва за предсказване на базата на зададени данни, като предсказанията се разделят на няколко вида: Classification, Regression, Clustering, Dimensionality reduction, Model selection, Preprocessing

CountVectorizer – Превръща данни от текст в числови данни. Като изход извежда колко пъти всяка дума е използвана в даден текст.





Картина, която съдържа текст

Описанието е генерирано автоматично

TfidfVectorizer – Превръща данни от текстов вид във векторен такъв, като може да се използва за вход на estimator.

Картина, която съдържа текст, тъмен, оранжев

Описанието е генерирано автоматично

cosine\_ similarity - Коефициент на подобност, получен като резултат сформиран от предварително зададена информация. В случая имаме функция, в която се въвежда заглавие на филм и cosine\_similarity и се извеждат 10 подобни филма според тази подобност.

Картина, която съдържа текст

Описанието е генерирано автоматично

pandas

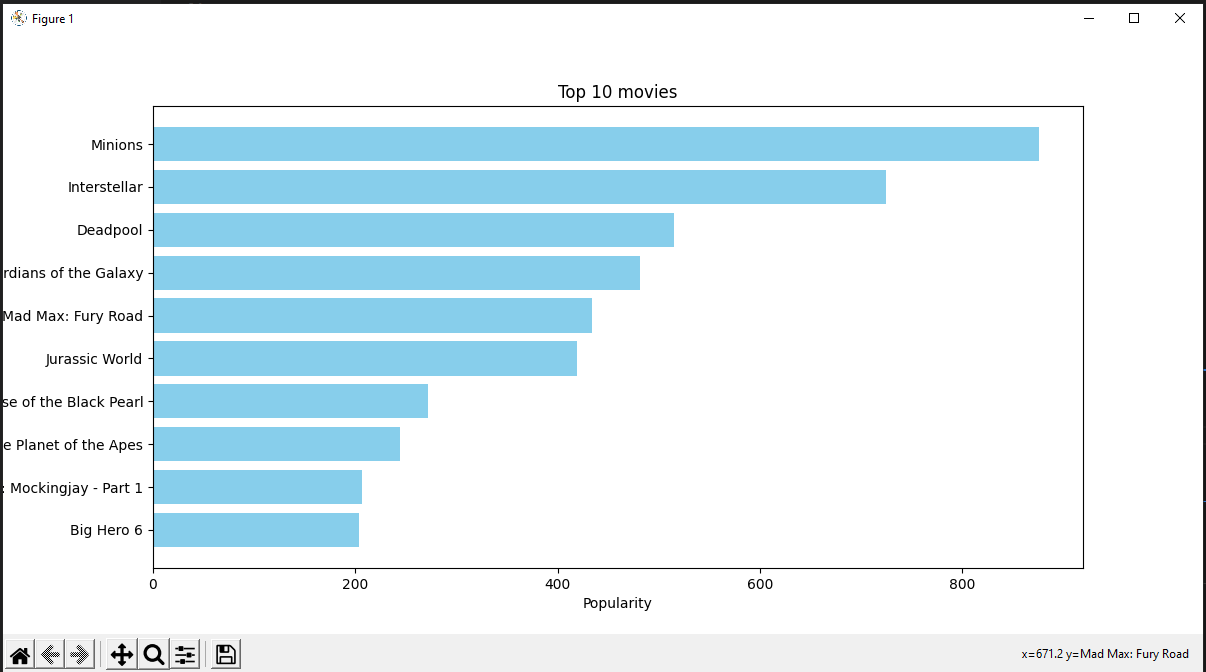
Използва се за анализиране и извадки на дата и обръщането им в pandas array.

Картина, която съдържа текст

Описанието е генерирано автоматично

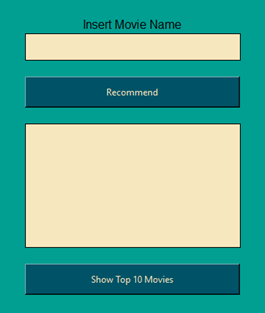
matplotlib

Библиотека за визуализация на данни



В този случай тя визуализира топ 10 филма, базирани на популярност от базата данни, която се използва. Засега базата данни е статична, следователно резултатите не търпят промяна.

Tkinter – Вградена библиотека в Python за изготвяне на графичен интерфейс.

Картина, която съдържа текст

Описанието е генерирано автоматично

literal\_eval от ast(Abstract Syntax Tree) е функция, която разчита файл или низ и определя от какъв тип е стойността, съдържаща се в тях. Низът или файлът могат да съдържат една от следните структури на Python: strings, bytes, numbers, tuples, lists, dicts, sets, booleans, and None.

В проекта признаците, по които се сравняват филмите са жанр, актьори, продуцент и ключови думи, тъй като чрез тях може да се проследи всяка прилика между няколко филма, включително тематика.

