What movie should I watch tonight

**INTRODUCTION**:

* A recommendation device is a type of statistics ltering device which attempts to are awaiting the options of a user and make suggests based on the one’s opportunities.
* There is a wide form of packages for advice structures. These have grown to be an increasing number of well-known over the previous couple of years and in the meantime are applied in maximum online systems that we use. The content of such structures varies from movies, song, books and motion photos, to buddies and testimonies on social media structures, to products on e-trade internet websites, to people on expert and dating websites, to look results back on Google. These systems are often used to collect information about user choices and moreover we can use this information for improving suggestions in future. We can take an example of Facebook Can monitor your interaction with various stories on your feed in order to learn what types of Stories appeal to you.
* There are so many applications for recommendation system which are becoming popular and are making use in online portal like movies, books, videos, meetings etc.
* For example, in case of television show collaborative filtering recommendation system will predict which shows user like to watch more based on the show’s TRP i.e. target rating point
* Every user has different likes and dislikes. Even the taste of single user can vary depending on many factors such as mood, season or type of activity user is doing.
* Sometimes, the recommender systems cam makes improvements based on the activities of large number of people. We can take an example of Amazon in which large number of customers buying apple MacBook who also buy USB-C-to-USBA Adaptor, they can recommend the Adaptor to a new user who had just added a MacBook to is cart.
* We used collaborative filtering to group similar user together and use information about the group to make recommendation to the user.
* **About the dataset –**

This dataset consists of following files: -

1. **Movies\_metadata.csv**: - This file contains an information about 45,000 movies features which includes posters, backdrops, budget, revenue, release dates, languages, production countries and companies.
2. **credits.csv**: Consists of Cast and Crew Information for all our movies. Available in the form of a stringified JSON Object.
3. **ratings\_small.csv**: The subset of 100,000 ratings from 700 users on 9,000 movies.

* **Types of filtering-**

1. Demographic filtering
2. Collaborative filtering
3. User based filtering
4. Item based filtering
5. Using SVD (SINGLE VALUE DECOMPOSITION)

* **How this topic came to our mind: -**

1. Different concept
2. More demand in the market
3. It makes predictions based on the taste of the users
4. It tackles the issue of data trouble from large amount of data or information.
5. Learnt many things on this topic
6. We can Learn different algorithms
7. Topic is quite good to work on as we can learn a lot from this as most of the big companies earn lot or fame with this type of projects

* **Which type of filtering was best among the 3: -**

1. Among all 3 SVD filtering is the best one as SVD deal with the scalability and sparsity issue created via CF is to leverage a latent thing version to capture the similarity among users and objects.
2. SVD decreases the dimension of the utility matrix by extracting its latent factors. Therefore, it helps us better understand the relationship between users and items as they become directly comparable.

**Related Work**

We can take related work example of **Netflix**.

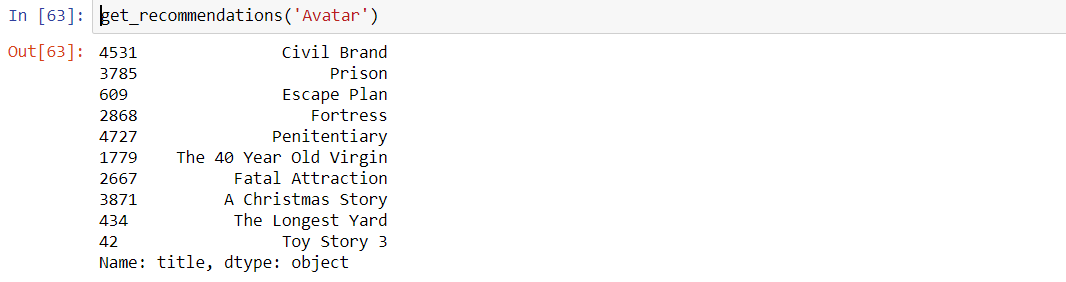
* Netflix is one of the biggest internet television networks with 158 million paying subscribers across the globe. Starting of 1997 people to rent DVD’s online in 2007 it came out with its own streaming services, customers could watch shows and movies on their computers, TV screens, phones, gaming devices. In 2013 it directly competing with TV networks and cable for original content, the reason behind Netflix is Machine learning, AI and big data analytics
* Netflix uses a content delivery network to store and transmit movies and TV shows depending on the locations, popularity and purchasing right Netflix display different content for its users that is why what you watch in Canada is different from what you watch in India
* How Netflix recommend these shows? Netflix uses its recommendation engine.
* Machine learning is necessary for this method it uses user data to informed suggestions later such as water your wash previously your rating, search history and so all this complex data analyzed and finally a conclusion is drawn.
* It uses algorithms to extract high quality images from the respective videos with the help of static visual analysis and creativity, Netflix uses various algorithms to decide its content and to make a personalized recommendation list.

**Methods**

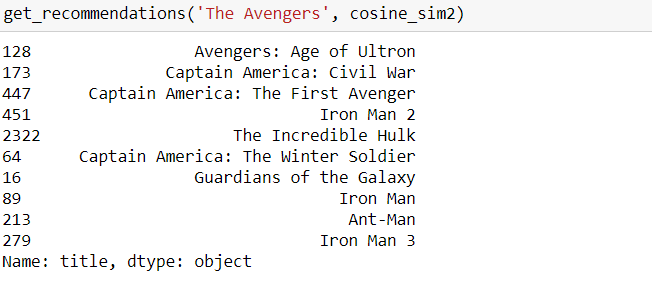
* + **ALGORITHMS**
* **Demographic Filtering**: - it is used basically such as identity of different types of users such as gender, education and certain items. It infers recommendation rules based on stereotypes.
* **Collaborative Filtering**: -
* **User based Filtering:**-Two types of method are used in user based filtering:-
* Cosine similarity: Cosine similarity is a metric used to measure how similar the documents are irrespective of their size. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space.
* Pearson coefficient: is the [covariance](https://en.wikipedia.org/wiki/Covariance) of the two variables divided by the product of their [standard deviations](https://en.wikipedia.org/wiki/Standard_deviations).
* **Item based filtering: -** There was some problem with the user-based system, so we built recommendation system that can recommend movies based on the actors, crews, keywords and genres. It goes without saying that the quality of our recommender would be increased with the usage of better metadata. We are going to build a recommender based on the following metadata: the 3 top actors, the director, related genres and the movie plot keywords
* **Single Value Decomposition**: - In linear algebra, the singular value decomposition (SVD) is a factorization of a real or complex matrix. It is the generalization of the eigen decomposition of a normal matrix (for example, a symmetric matrix with non-negative eigenvalues) to any matrix via an extension of the polar decomposition.

**Results:-**

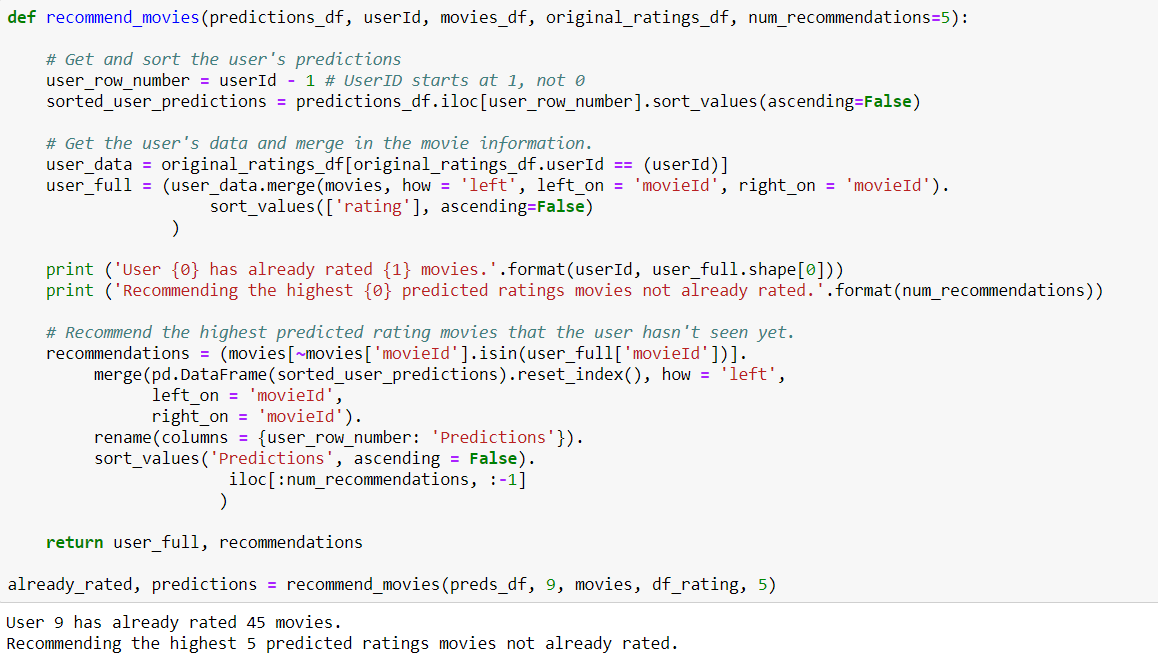
Demographic filtering recommendation

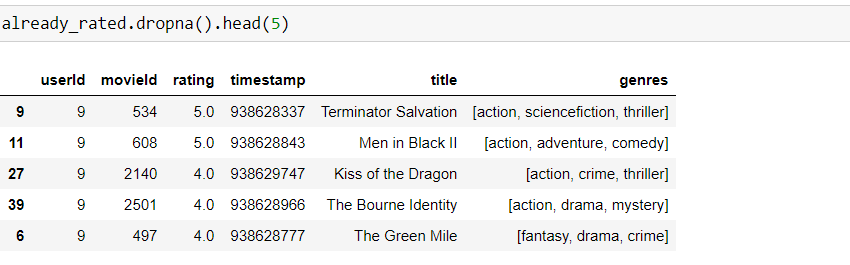


User Based Filtering

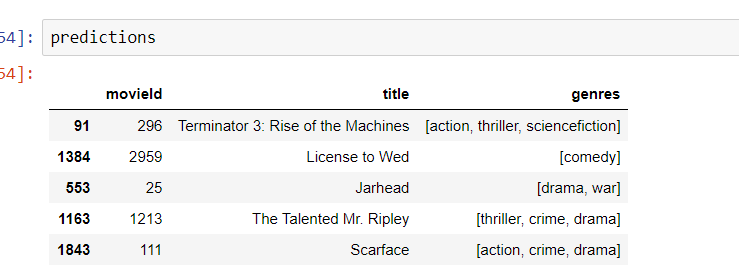


Item Based Filtering





Collaborative Filtering Using SVD



Predictions based on SVD

We tried 4 methods for our recommendation engine. The method which worked best out of all was Collaborative Filtering using SVD because it gave us better understand the relationship between users and items as they become directly comparable . Also, the recommendation worked well for the users as it gave almost similar predictions as recommended and rated by the user

**Discussion: -**

* We found so many roadblocks in this project but not any failure.
* There were so many challenges we confronted while making the project like firstly finding the dataset to get the best output that matches with our collaborative filtering technique was the greatest test.
* What all libraries to import and after importing it gave lots of errors
* Collaborative filtering was completely new for us and applying the new technique in the project was bit challenging
* Shared the ideas what to do further in the project
* When doing filtering with SVD we encounter many errors while trying different libraries and it became very difficult for us that how we can try implementing the last step of our project. Thanks to our sir who helped us found library for svd with which we were able to solve the problem and implement it successfully.

**Conclusion:-** We basically made a recommendation engine which helped users in recommending movies based on the ratings,genre,content. Various algorithms were implemented and the one algorithm which worked well as compared to others was Collaborative filtering using SVD as it overcomes the shortcomings of the other filtering methods because it generates a relationship between item and user both.

**Contributions:-**

* Gurmanjit Singh Sahni: Found the dataset and worked on the codes and made plans regularly what to do every week in the project
* Arvind Sharma: explained the concept of user and item-based filtering and assisted gurmanjit singh in coding part as well
* Anchal Gupta: Described the notion of collaborative filtering and assisted gurmanjit singh in coding

**References:-**

* <https://www.kaggle.com/rounakbanik/the-movies-dataset-This> is the link for our dataset
* <https://www.liulangmao.org/?tag=python-> This was the most important link as this helped us what functions to use and where to use and the libraries required for importing.
* [https://medium.com/@cfpinela/recommender-systems-user-based-and-item-based-collaborative-filtering-5d5f375a127f- This](https://medium.com/@cfpinela/recommender-systems-user-based-and-item-based-collaborative-filtering-5d5f375a127f-%20%20This) helped us understand the concepts of user based and item based filtering.
* <https://stackoverflow.com/questions/1838806/euclidean-distance-vs-pearson-correlation-vs-cosine-similarity-> For understanding the concepts of cosine similarity and Pearson coeff.
* <https://heartbeat.fritz.ai/recommender-systems-with-python-part-iii-collaborative-filtering-singular-value-decomposition-5b5dcb3f242b-> This helped us in coding of svd.
* <https://hackernoon.com/introduction-to-recommender-system-part-1-collaborative-filtering-singular-value-decomposition-44c9659c5e75-> This helped us understand the single value decomposition

**Appendices:-**

* [data overview. ipynb](http://localhost:8892/notebooks/Desktop/DAB300%20labs/data%20overview.ipynb)
* [Data cleaning and exploratory data analysis. ipynb](http://localhost:8892/notebooks/Desktop/DAB300%20labs/Data%20cleaning%20and%20exploratory%20data%20analysis.ipynb)
* [Demographic Filtering.ipynb](http://localhost:8892/notebooks/Desktop/DAB300%20labs/Demographic%20Filtering.ipynb)
* [Collaborative filtering user based, item based and using svd.ipynb](http://localhost:8892/notebooks/Desktop/DAB300%20labs/Collaborative%20filtering%20user%20based%2C%20item%20based%20and%20using%20svd.ipynb)