Movie Recommender System

The primary goal of the project was to develop Movie Recommender System. However before that could have been possible development of Data Science Pipeline turn out to be needed too.

In a regard to the set of additional challanges the final strategy involved following objectives:

- scrapping the data from web,
- storing the data as RDMS,
- building views for data exploration and data visualisation,
- development of few classifiers.

To scrap the data I have used API service provided by themoviedb.org[1]. Desiging scrappers of iterative nature proved to be very helpful since each query returned single JSON document. The collections of documents were appended and saved under following data sets: Movies, Genres, Production Countries, Keywords, Cast, and Crew[2]. Predefined SQLite views were used to clean the outputs once each new batch of data entries was appended to already stored data. Hence to above the data sets used for later data processing - including multiple joins - had only unique rows. Once the amount of downloaded data turned to be large enough I went through Exploratory Data Analysis to famillliarize myself with information collected[3].

Although development of classifiers was not expected or used by recommender engines the part is added to exhaust the list of steps needed for Data Science Pipeline. Decission trees, logistic regression, neural networks, and XGBOOST classifiers were provided mainly to enhance exploratory part of the project[4].

Content Based Recommenders[5] and Collaborative filtering[6] are perceived as two major methods for building recommender systems. Although both are used and assessed in the project the combination of the two called Hybrid method[7] was not successfuly explored due to lack of data.

[1] More information related to the API can be found under following link: www.themoviedb.org/documentation/api .
[2] More information related to scrapper logic or output column names can be found in SCRAPPERS notebook.
[3] The details on EDA can be found in EXPLORATORY DATA ANALYSIS notebook.
[4] More information related to the classifires can be found under CLASSIFIERS notebook.
[5] https://en.wikipedia.org/wiki/Recommender_system#Content-based_filtering
[6] https://en.wikipedia.org/wiki/Collaborative_filtering
[7] https://en.wikipedia.org/wiki/Recommender_system#Hybrid_recommender_systems