

Capstone Project 1 Proposal

Problem Statement

Nowadays, we need recommender systems almost everywhere in our lives. Therefore, retailers are becoming more interested in recommender systems to analyze patterns of user interest in products and provide personalized recommendations. The first goal of this project is understanding, analyzing, and correlating the trend in average rating movies of different genres from 1995-2015. The second goal is building a recommender engine to predict the rating of each movie.

Business Objective

The recommender system is useful to any business that makes money via recommendations. Since we want to work with movie dataset, the client of this project could be Amazon, Netflix, Hulu, HBO, etc.

Giving good recommendations will help users spend less time searching for their type of movie and having a recommender platform. This will help the customer to continue with the service and having a good experience.

Data

The data for this project has been collected from a free, noncommercial movie recommender [MoviLens.org](https://grouplens.org/datasets/movielens). (<https://grouplens.org/datasets/movielens>) This recommender is like Netflix, minus the ability to watch movies. The dataset which was released in April 2015 contains 20 million ratings and 465,000 tag applications applied to 27,000 movies by 138,000 users (movies.csv, ratings.csv, link.csv). The data has been created in October 2016. Users were selected at random for inclusion. All selected users had rated at least 20 movies.

Solution

- 1- How to get the insight about how people's rating change for different genres over the years?
Getting different distributions of the movie-ratings and movie-relevancy across title, genre, year, and tag of each movie can give the insight about how people's taste vary over the years. Different inferential statistics and visualization techniques can be implemented to deliver a better understanding of the trends.
- 2- How to predict the rating?
In this work Collaborative filtering (CF) techniques will be used to analyze relationships between users and interdependencies among products to identify new user-item associations. Most CF algorithms are based on a user-item rating matrix where each row represents a user, each column an item. The entries of this matrix are ratings given by users to items. One of the primary areas of collaborative filtering is the latent factor models which

are based on matrix factorization. In this work, the Singular Value Decomposition (SVD) technique will be used to predict the ratings.

Workflow

- 1- Collecting data and applying data wrangling methods
- 2- Starting exploratory data analysis to find trends
- 3- Conduct further data analysis to identify relationships between different variables
- 4- Perform in-depth analysis, using machine learning or another advanced technique to predict

Deliverables

- Final report
- Jupyter notebook .ipynb code
- Slide deck
- Presenting the project