In this assignment, we are tasked to implement a movie recommender system that Is able to learn from the user’s movie view history and ratings of movie they have watched and predict what the user would give to a movie that has not yet been rated. In order to build the movie recommender system, we are provided with data on users and movies. The dataset include information on actors, Directors, ratings, generes, and movies. To build the movie recommender system, we have opted on a content-based filtering approach where the algorithm recommends products which are similar to the ones the user has liked in the past.

Pre-processing

Since the dataset comes from different files and tables, we started by first filtering and converting the dataset into a single pands data frame. we assembled all the movie profiles (the generes, movie tags, actors, directors, ratings) and indexed the table by the movie ID so it will be easier for us to retrieve information later in the classfication step. One thing we have noticed is, since all the data comes from different sources, there are missing information on some features when we tried to merge all the data into one. This could have an impact on our metrics. So, to counteract this effect we wanted to give weight to each feature so missing information can’t be given a default value.