Team member: Daniela Florit
Title: IMDB movie rating prediction:

Summary:

The objective of this project is to build a classifier capable of predicting the IMDB ratings of movies. The possible labels are integers from 0 to 10 (inclusive). The dataset used for the learning algorithm will be the one provided in http://Kaggle.com (link provided below). This dataset consists of 5043 movies and 28 variables per movie. One of the challenges, other than building a strong learning algorithm, is to build a good feature extractor that takes into consideration discrepancies on the dataset, such as for example, the amount of money the movie made, in which case the same number can be interpreted differently depending on the year. This problem is especially interesting due to the massive interest of the public in movies and the use that companies such as Netflix can give to classifiers such as this one. It would also be interesting to know which of the features in the feature vector is the one that contributes the most to the rating of the movie. This will provide us with some insight into what makes a movie good (the budget, the director, the actor, etc). This data can also be used to maximize the commercial value of movies that are currently being produced. Additionally, finding what features contribute the most to the rating would be useful during the marketing of the movie, as these features could be considered the selling points of the movie, and therefore could be highlighted as the movie is commercialized.

Database source: https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset