Final project proposal

Debora Fieberg, Max Simon

January 7, 2018

1 PROJECT 1: FLIGHTS

2 PROJECT 2: MOVIES

In this project we want to investigate a dataset on movies which can be found on Kaggle[1]. This dataset contains about 329000 instances with features like the film title, genre, release date, popularity and votes.

In our investigation we want to analyse the different titles and how the correlate with the success of the movie and how this changed over time in the different genres.

To this end, we will preprocess the titles with common techniques like stemming or lematizing and construct different features from it (e.g. word bag, first word of title, etc.). Afterwards we will analyze important key words and their influence on the movie success. Also an automatic title creation from the learned titles might be possible.

---

[1] https://www.kaggle.com/stephanerappeneau/350-000-movies-from-themoviedborg, 07.01.18, 15:00