

Machine Learning

*Master of Arts in Banking and Finance*

*Spring term 2020*

**Development of a movie recommendation engine**

Group term paper

Nils Kaufmann | nils.kaufmann@student.unisg.ch | 19-602-796

Yifan Zhu | yifan.zhu@student.unisg.ch | 19-600-105

Jan Scheidegger | jan.scheidegger@student.unisg.ch | 15-055-411

Daniel Rüger | daniel.rueger@student.unisg.ch | 09-593-377

submitted on 14 May 2020

Prof. Dr. Anna-Lena Horlemann

**Abstract**

Today, platforms such as Netflix or Amazon offer tens of thousands of films which can be streamed online. With this huge amount, the user faces an information overload problem which makes the choice of a movie best-suited to their interests and needs time-consuming and complicated. In order to increase convenience and the quality of movie selection, we present three approaches for a moving recommendation system using machine learning algorithms. [To be finished…]

**List of figures**

[tbd]

# Introduction

The data available on the internet is constantly and quickly increasing (Bridle, 2010, p. 4). Consumers who inform themselves using the internet are confronted with a large amount of information which can easily become overstraining (Bawden & Robinson, 2009, p. 183). Information overload can result (Cui, 2017, p. 1). Consumers need to invest considerable time and resources in order to analyse information, and choosing a product can become complicated (Chan as cited in Li, 2016, p. 1; Keselmann, Rosemblat, & Kiligoclu as cited in Li, 2016, p. 1). A field which is affected by this evolution is movie selection (Cui, 2017, p. 1). Platforms such as Netflix or Amazon offer a huge number of films which customers can stream online (Peng, Liangshan, & Xiuran, 2013, p. 1). Thus, for consumers, it is difficult to identify the product best-suited to their needs.

Information overload can be reduced through recommendation systems, which screen information. These algorithms allow users to individualise recommendations with respect to consumer characteristics such as personal taste and needs. Recommendation systems have been successfully implemented in order to improve recommendation results, such as in the field of movies (Wei, Zheng, Chen, & Chen, 2016), music (Mao, Chen, Hu, and Zhang, 2016), and news (Wang & Shang, 2015). Netflix, for example, considers the films a client has watched and suggests further movies which are comparable to them (Reddy, Nalluri, Kunisetti, Ashok, & Venaktesh, 2019, p. 392).

Various methods can be used to implement a recommendation system. Three important categories are content-based, collaborative, and hybrid algorithms. Content-based systems analyse the user’s past consumption behaviour and suggest movies based on this. Collaborative filtering examines the past ratings and experiences of consumers and compares this with other clients’ ratings and experiences. Based on the choices of the users being the most comparable, suggestions are generated. In order to avoid the drawbacks of both methods, hybrid models have been proposed. (Reddy et al., 2019, p. 392)

The aim of this paper is to develop a movie recommendation system using machine learning algorithms. Our product will help consumers save time and resources when choosing a movie, make the process more convenient, and help consumers find films better-suited to their tastes and needs.

We propose three different approaches for overcoming the disadvantages of each of the above methods. We first implemented a demographic filtering algorithm, then a content-based algorithm, and finally several collaborative filtering algorithms. The demographic filtering algorithm allows to suggest films for persons with unknown characteristics. The content-based approach uses a k-NN algorithm and a plot-based recommender. Both use the consumer history of the user. In order to compare the experiences of different consumers, we used a collaborative filtering algorithm. [Finally, … tbd]

There are different datasets for movies that can be used to develop a movie recommendation system. We have chosen The Movie Database (TMDB) 5000 movie dataset from the online community Kaggle because it contains many variables such as revenues and ratings. The many variables allow for a large variety of machine learning algorithms to be developed. TMDB 5000 consists of 4808 individual movies. An alternative dataset would have been the Internet Movie Database (IMDB) dataset. However, this dataset has over 40 characteristics for every individual film, making cleaning and preparing the data complicated. (https://www.kaggle.com)

The subsequent parts of this paper are structured as follows. First, we describe the dataset in more detail. In the application section that follows, we describe the algorithms we used and their actual implementation. Finally, we end with the results and conclusion.

Due to the limited scope of this article, we do not test all the predictions made by our recommendation engine. Again due to the limited scope, we do not determine how the three approaches can be used together in a single recommendation. Instead, the user should assess the three individual algorithms on their own.

Many papers which present movie recommendation systems using machine learning methods have been published in recent years. A k-NN collaborative filtering algorithm, for instance, was used by Cui (2017). Reddy et al. (2019) applied a content-based method which used genre correlations. Wang, Sang, Zeng, and Hirokawa (2017) implemented a support vector machine method and improved particle swarm optimisation.