

Recommender Systems: Implementation and Evaluation

Aindrila Basu
201570162

Supervised by Daniel Garcia (Kainos Group), Dr. Diego Capozzi (Kainos Group),
Liam Coyle (Kainos Group) and Dr. Luisa Cutillo (University of Leeds)

Submitted in accordance with the requirements for the
module MATH5872M: Dissertation in Data Science and Analytics
as part of the degree of

Master of Science in Data Science and Analytics

The University of Leeds, School of Mathematics

September 2022

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Abstract

Recommender Systems have been addressing the issue of information overload for over two decades now. With the advent of internet and e-commerce, they have become a necessity for companies that want to survive in this digital era. There are different kinds of recommender systems, following different strategies that are in use today. Industry giants like Netflix, Amazon, Google all employ recommender systems in their websites and applications.

The aim of this project is to develop and evaluate three types of Recommender Systems on a publicly available dataset (1.37 million customers) as a Minimum Viable Product for a renowned fashion retailer. These three types are: Non-Personalised Recommender System, Content-Based Recommender System and Collaborative Filtering using Matrix Factorisation (ALS) Recommender System. The development of the MVP is embedded into the retailer's business principles to enable the application of marketing strategies. The algorithms for the three techniques were developed and implemented, and finally evaluated with three different metrics, namely Mean Precision at k, Mean Recall and Mean Average Precision at k. A comparative study of the three types was performed on three different numbers (10, 50 and 100) of recommendations generated.

The results of the comparison among the three techniques enabled the identification of the Non-Personalised Recommender System as the best performing for our dataset. In terms of MAP@k, this system performed 3.3 times and 2.5 times better than the Content-Based and the ALS ones respectively. This recommender system can then be integrated with the marketing principles for achieving higher revenue generation.

Several strategies can be used to improve the quality of the recommendations generated. Machine Learning techniques such as Natural Language Processing (NLP) could be used to find similar products. Convolutional Neural Networks (CNN) could also be used for feature extraction from images to enhance the performance of the recommender systems. All of these could be explored in the future.