Movie Recommendation System

There are many approaches towards creating Movie Recommendation System. Our research investigated the idea of utilizing **K-means algorithm** in order to offer personalized movie suggestions for each user.

Introduction

Traditional movie recommendation systems often rely on single perspective of aggregating the preferences of the entire user base. This approach struggles to entail the nuances of unique movie tastes of the individual users. To address this issue, our research explored the idea of a more personalized recommendations...

Objective

Research the viability of K-means clustering approach as a means to offer personalized recommendations to the users and increase the accuracy of user's opinion on the movie.

Methodology

We used **MovieLense** Latest Datasets, to gather information about users and their movie preferences. Then using K-means algorithm we grouped the users into **clusters based on the similar genre preferences**. Lastly to predict how a user would rate a particular movie, we calculated the mean rating of said movie of all users in the same cluster

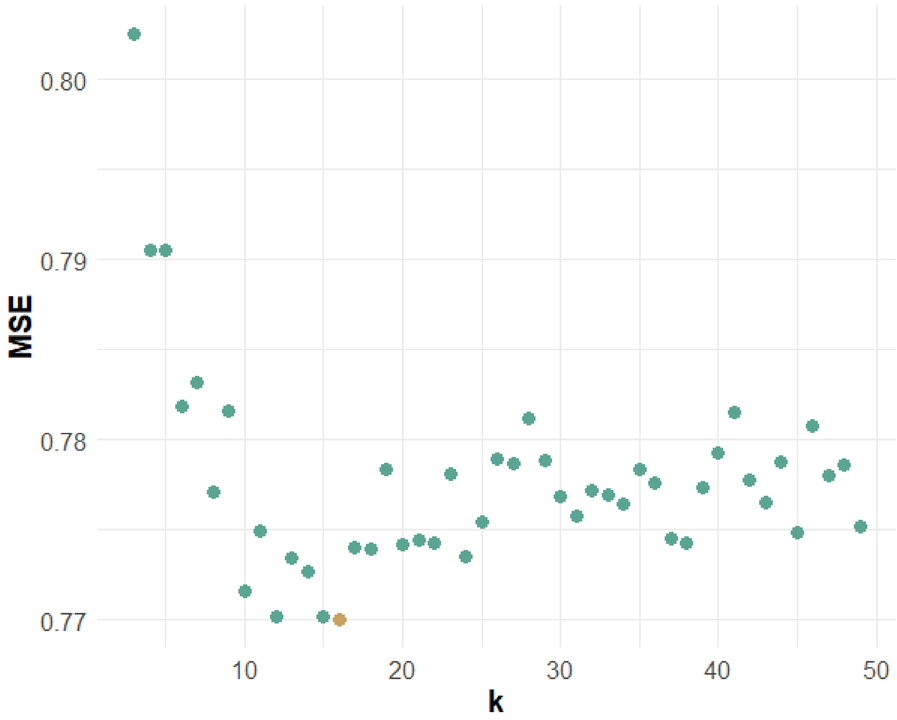
Results

Generic prediction based on the mean score of all users (k=1) produced a MSE of 0.97. Increasing the number of clusters increased the accuracy of the prediction. The best number of clusters for the data **k=16**, allowed us to achieve **MSE** of 0.77

Analysis

We skipped k=1 and k=2 from the plot as to visualize more precisely the differences in MSE between clusters.

At first when the number of clusters increases the MSE score plummets, reaching it lowest point for k=16. Past that point, for k>16, MSE value seems to be on a slow but steady increase. Whilst still being better than the baseline, high number of clusters would not be that beneficial to us.



MSE value for each number of clusters k from the range [3,50]

Conclusion

Our research **proved** that utilizing K-means clustering algorithm results in more personalized recommendations and **better prediction accuracy**, than recommendations based on the whole database.

Authors: Maciej Janicki, Marcin Kapiszewski, Marcel Rojewski, Adam Tomys

POZNAN UNIVERSITY OF TECHNOLOGY