MovieOcean: Assessment of a Personality-based Recommender System

Luca Rolshoven[1], Corina Masanti[1], Jhonny Pincay[2,3][a], Luis Terán[2], José Mancera[2]
and Edy Portmann[2][b]

[1]University of Bern, Hochschulstrasse 6, Bern, Switzerland
[2]Human-IST Institute, University of Fribourg, Boulevard de Pérolles 90, Fribourg, Switzerland
[3]Pontificia Universidad Católica del Ecuador, Av. 12 de Octubre 1076, Quito, Ecuador

Keywords: Recommender Systems, Personality-based Recommenders, Personalized Recommendations, Big Five Model.

Abstract: This research effort explores the incorporation of personality treats into user-user collaborative filtering algorithms. To explore the performance of such a method, MovieOcean, a movie recommender system that uses a questionnaire based on the Big Five model to generate personality profiles, was implemented. These personality profiles are used to precompute personality-based neighborhoods, which are then used to predict movie ratings and generate recommendations. In an offline analysis, the root mean square error metric is computed to analyze the accuracy of the predicted ratings and the F1-score to assess the relevance of the recommendations for the personality-based and a standard-rating-based approach. The obtained results showed that the root mean square error of the personality-based recommender system improves when the personality has a higher weight than the information about the user ratings. A subsequent t-test was conducted for the proposed personality-based approach underperformed based on the root mean square error metric. Furthermore, interviews with users suggested that including aspects of personality when computing recommendations is well-perceived and can indeed help improve current recommendation methods.

SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

While collaborative filtering methods are widely used in implementing recommender systems (RS) with good results, there is still room for improvement in the quality of recommendations. The advent of streaming platforms in the last few years (e.g., Netflix, Disney+, and Paramount+) and the easiness of accessing a broad amount of media content have led many companies and researchers to find methods that provide better recommendations to engage users. Because personality is about thinking, feeling, and behaving, it seems coherent to think that it is possible to improve the utility of recommendations by harnessing personality data.

Authors such as Aaker (1999) had found that brands that were associated with a set of personality traits were perceived as more favorable when the individuals were schematic on the personality dimension that was also highly descriptive for the brand. Mulyanegara et al. (2009) employed the Big Five personality scale to assess the personality of 251 subjects in five dimensions and used those assessments to measure the relationship between consumer personality and brand personality in the context of fashion products. Although the results may differ in each domain, they suggest that including personality aspects in the implementation of RS engines could enhance the quality of recommendations. The results of the research work of Weaver Weaver III (1991) support the conception that there is a correlation between personality traits and media preferences across a variety of media.

This work addresses the following questions: (i) how reliable are personality-based neighborhoods for the computation of recommendations, and (ii) to what extent are personality-based recommendations perceived as more valuable than those based on collaborative filtering? To answer these questions, the authors implemented *MovieOcean*, a personality-based movie recommender system that uses the Big Five model to generate personality profiles for each user. These profiles are used to

[a] https://orcid.org/0000-0003-2045-8820
[b] https://orcid.org/0000-0001-6448-1139

Rolshoven, L., Masanti, C., Pincay, J., Terán, L., Mancera, J. and Portmann, E.
MovieOcean: Assessment of a Personality-based Recommender System.
DOI: 10.5220/0011002500003179
In *Proceedings of the 24th International Conference on Enterprise Information Systems (ICEIS 2022) - Volume 1*, pages 690-698
ISBN: 978-989-758-569-2; ISSN: 2184-4992
Copyright © 2022 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

precompute personality-based neighborhoods, then used to predict movie ratings and generate recommendations. Furthermore, the resulting recommendations are assessed by conducting experiments and measuring the root mean square error (RMSE) and F1-score.

This article is structured as follows: Section 2 presents the theoretical background in which this research effort is grounded. The methods developed and applied in the execution of this work are described in Section 3. Results are presented in Section 4. Section 5 finishes the article with a summary of the work conducted and the lessons learned.

2 THEORETICAL BACKGROUND

Through the presentation of related work, this section introduces the concepts and theories used to address the above mentioned research questions.

2.1 Personality and the Big Five Model

The American Psychological Association (APA) defines personality as follows (American Psychological Association, 2020): *"Personality refers to individual differences in characteristic patterns of thinking, feeling, and behaving. The study of personality focuses on two broad areas: Understanding individual differences in particular personality characteristics, such as sociability or irritability. The other is understanding how the various parts of a person come together as a whole."*

When discussing personality and personality characteristics in differential psychology, it is essential to provide a shared taxonomy. The Big Five personality model was developed as a result of the findings of many study initiatives. It is made up of five components (John et al., 2008):

- **Openness:** People with high openness tend to be more original, open-minded, experimental, and creative.
- **Conscientiousness:** People with high conscientiousness tend to be more orderly, responsible, goal-oriented, and organized.
- **Extraversion:** People with high extroversion tend to be more energetic, enthusiastic, active, sociable, and assertive.
- **Agreeableness:** People with high agreeableness tend to be more altruistic, affectionate, modest, trustful, and cooperative.
- **Neuroticism:** People with high neuroticism tend to be more nervous, anxious, sad, and tense.

Which can be abbreviated as *OCEAN*.

The *Big* in Big Five was chosen to emphasize the fact that these five factors are extensive and that they were selected to represent personality at the broadest level of abstraction (Pervin and John, 1999).

2.2 Personality Correlations

People differ significantly from each other based on their personality traits. In psychology, researchers aim to find the most defining characteristics detailed enough to capture a personality but general enough to avoid high complexity. Various studies have proven a strong correlation and have also shown how it can produce relevant predictions.

For instance, Rentfrow and Gosling (2003) demonstrated how personality traits influence the preferred music type. Also, Barrick and Mount (1991) found out that there is a close relationship between personality characteristics and job performance. Other studies showed how personality could influence the everyday online behavior of a person. Preferences for specific social networks or the amount of time that one spends online are also correlated with the personality of the user, as Zhong et al. (2011) were able to present in their manuscript. Besides, Hu and Pu (2013) managed to show the correlation between personality characteristics and user rating behavior for retail products.

Cantador et al. (2013) looked at the relationship between personality types and tastes in different entertainment fields. Movies, TV shows, music, and books were among the domains they selected. They looked at data from 53 226 Facebook users to see any significant links between personality traits and preferences in the domains listed. In each of the chosen domains, they generated personality-based user stereotypes and association rules for 16 genres.

2.3 Other Related Work

In the literature, it is possible to find several approaches to personality-based recommender systems. Paiva et al. (2017) proposed a hybrid recommendation approach that combines the Big Five personality traits of the users with a correlation between car fronts and power and sociability perceptions to be used for semantic searches in vehicle sales portals. They used the BFI-10 inventory (Rammstedt and John, 2007), which solely consists of 10 questions, to assess the Big Five traits of their users. Then, they used the Euclidean distance between users to form a neighborhood of size $k = 3$, which is used with the correlations described above to offer recommendations.

Braunhofer et al. (2015) used the Five-Item Personality Inventory (FIPI), which includes even fewer questions, to assess the personalities of the users and then incorporate them into their recommender system for Point of Interest (POIs) recommendations. Their recommendation algorithm makes use of matrix factorization (Koren et al., 2009) and incorporates the demographic data of the users as well as their personalities.

Another approach is to use a standard user-user collaborative filtering algorithm and change the similarity metrics to include personality information. Tkalčič et al. (2009) used this method and determined that their personality-based similarities led to an F1-score that was statistically equivalent to the one obtained by using rating-based similarities. However, they state that the personality-based approach has three advantages: (i) an initial questionnaire solves the new user problem, (ii) the similarity computation between users is less expensive, and (iii) the impact of the sparsity problem is lowered. To assess the personalities of the users, the authors used a version of the International Personality Item Pool (IPIP) with 50 questions (Goldberg et al., 2006).

In the realm of movies, Nalmpantis and Tjortjis (2017) investigated the impact of integrating personality into the recommendation process. They used a hybrid approach to compute the recommendations, in which a questionnaire was used to measure the user's personality characteristics. The personality scores are then used together with the findings of Cantador et al. (2013) to identify genres that the consumer might enjoy. They estimated movie ratings using collaborative filtering, for which users had to rate 20 movies in advance. Finally, the recommendations were generated by considering the movie ratings and the user's personality with a weight of 0.5 each. They compared their system to one that only considers recommendations based on ratings and discovered that users preferred the 50%-50% approach to the rating-based method.

Considering previous research findings, using personality characteristics to perform movie suggestions seems to be a suitable way of enhancing the quality and usefulness of the recommendations. Compared to previous efforts, this work aims to evaluate the effect of creating a personality-based neighborhood and determine whether incorporating personality traits increases the perceived value of the recommendations among users.

3 METHOD AND USE CASE

The approach used in this study consists of three major stages: (i) personality assessment, (ii) recommendation engine design, and (iii) evaluation. The following sections go into greater detail about these phases and the implementation of MovieOcean. Besides that, the movie information was obtained from The Movie Database (TMBD)[1] through its free developer API.

3.1 Personality Assessment

The personality of a user was determined with the help of the Big Five model introduced in Section 2.1. A standardized questionnaire containing 50 questions[2] from the (IPIP) was used. This questionnaire is managed by the Oregon Research Institute (Goldberg et al., 2006).

When the questionnaire is applied, each participant can score up to 50 points in each personality dimension (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). The higher the score in a dimension, the more the person can be described by this specific personality trait. The particular questions consist of statements that have to be rated on a 5-point Likert scale from *Very Inaccurate* to *Very Accurate*. There were precisely ten questions for each personality dimension, some of them being *positively keyed* and some of them being *negatively keyed*. For a question that is positively keyed, the answer *Very Inaccurate* would add one point to the corresponding personality trait. In contrast, the answer *Very Accurate* would result in five points added to that trait. For questions that were negative keyed, the opposite order applies.

After completing the questionnaire, users could see how many points they had obtained in each personality dimension.

3.2 Recommendation Engine Design

To build the recommendation engine, a user-user collaborative filtering algorithm was used. Neighborhoods of size five were built for each user. Neighborhoods were recalculated each time a new user completed the IPIP questionnaire. Recomputation of the neighborhoods was also triggered if a user with an existing personality profile decided to change one or more of their answers in the questionnaire. Equation 1 was used to calculate the similarity of two users a and k.

[1] https://www.themoviedb.org/documentation/api

[2] https://ipip.ori.org/new_ipip-50-item-scale.htm

$$\mathrm{sim}(a,k) = \frac{1}{1+p}\left[\mathrm{sim}_{\cos}(a,k) + p \cdot \mathrm{sim}_{\mathrm{pers}}(a,k)\right] \tag{1}$$

The variable p is the personality factor used to change the importance we want to give to the personality in contrast to the ratings. In MovieOcean, the value of p was set to 2, as the main focus was on the personality more than on the ratings. Equation 2 was used for calculating the cosine similarity between users a and k.

$$\mathrm{sim}_{\cos}(a,k) = \frac{\sum_{i \in I} r_{k,i} \cdot r_{a,i}}{\sqrt{\sum_{i \in I} (r_{k,i})^2} \cdot \sqrt{\sum_{i \in I} (r_{a,i})^2}}, \tag{2}$$

where I is the set of movies that were rated by both users and $r_{k,i}$ is the rating of user k for movie i. The personality similarity $\mathrm{sim}_{\mathrm{pers}}(a,k)$ of the two users can be calculated by using Equation 3:

$$\mathrm{sim}_{\mathrm{pers}}(a,k) = \frac{1}{\|\phi(a) - \phi(k)\| + \varepsilon} \tag{3}$$

In Equation 3, $\phi(a)$ corresponds to the personality embedding, (i.e., a vector containing the scores for each personality dimension of user a). Moreover, the term ε is a small value used to avoid a division by zero. The predicted rating of a movie i for user a is then calculated using the standard prediction formula in Equation 4.

$$r_{a,i} = \bar{r}_a + \frac{\sum_{u \in N_a} (r_{u,i} - \bar{r}_u) \cdot \mathrm{sim}(a,u)}{\sum_{u \in N_a} |\mathrm{sim}(a,u)|} \tag{4}$$

In the formula above, N_a is a set of users in the neighborhood of the active user a, $\bar{r}_a$ is the mean of all the ratings issued by user a and $\bar{r}_u$ the mean rating of user u respectively. This approach works as long as some users already have an established personality profile and ratings in the system. Otherwise, it cannot deliver good results. This issue is known as the cold start or new user problem. To address this problem, movies with high ratings that were of a genre well received by this particular personality type were recommended to the user. To establish the link between the personality and movie genres, the results of Cantador et al. (2013) were used. Then the three movie genres that had the least Euclidean distance to the user's personality were selected. The user would then receive recommendations based on these three genres and existing movie ratings, fetched from TMDB.

3.3 Evaluation Method

The personality-based approach presented in this work was compared with the standard rating-based approach of user-user collaborative filtering to evaluate the recommendation engine. Experiments with different similarity functions and various values for the neighborhood size and the personality factor were conducted. Besides, it was verified how the results changed by considering additional weighted demographic data. To numerically capture which approach performs better, two measurements were applied: RMSE for assessing the accuracy of the estimated ratings and the F1-score to measure the relevance of the recommendations.

3.3.1 Similarity and Neighborhood

Experiments with different similarity functions and different values for the neighborhood size were conducted to find the best possible combination. For the similarity function, four different methods were chosen Choi and Suh (2013): the cosine similarity, the Pearson correlation coefficient (PCC), the inverted Euclidean distance, and the Jaccard index. Our grid search procedure looked at neighborhoods of sizes 5, 7, 9, 11, 13, and 15. In practice, the neighborhood size is often larger than 15. In this case, however, large neighborhoods were not used because the dataset was small. A neighborhood of fifteen users is already big compared to the total number of users that we could include in the evaluation. We first used the similarity function to build the neighborhoods and then calculated the predicted ratings of each user. In each step of the grid search, one of the four similarity metrics was used to build the neighborhood, either in the space of user ratings or in the space of user personalities. However, there was one exception: when looking at the Jaccard Index, it was not used to build the personality-based neighborhoods but only to compute the rating similarity, which was still considered in the personality-based predictions of the ratings. Instead, the PCC was used to compute the neighborhoods for the personality-based approach. The Jaccard Index was not used to compute the neighborhood of similar personalities because it does not make sense in this scenario since the Jaccard Index would always yield 1.

The following equations were used to compute the predicted ratings, which are a modified version of the standard collaborative filtering algorithm (McLaughlin and Herlocker, 2004):

Rating: $\mathrm{sim}(a,k) =$

$$\frac{1}{1+d}\left[\mathrm{sim}_{\mathrm{rating}}(a,k) + d \cdot \mathrm{sim}_{\mathrm{dem}}(a,k)\right] \quad (5)$$

Personality:

$$\mathrm{sim}(a,k) = \frac{1}{1+p+d}$$
$$\left[\mathrm{sim}_{\mathrm{rating}}(a,k) + p \cdot \mathrm{sim}_{\mathrm{pers}}(a,k) + d \cdot \mathrm{sim}_{\mathrm{dem}}(a,k)\right] \quad (6)$$

In these equations, $\mathrm{sim}_{\mathrm{rating}}$ corresponds to the rating similarity between two users, $\mathrm{sim}_{\mathrm{pers}}$ to the similarity between the personalities of two users and $\mathrm{sim}_{\mathrm{dem}}$ is the demographic similarity between the users. The factors p and d correspond to the *personality factor* and *demographic factor*. They are used as a weight to give importance to the personality or demographic data. In the first experiment, $p = d = 1$ was set, and only the different similarity metrics and neighborhood sizes were changed. For $\mathrm{sim}_{\mathrm{rating}}$ and $\mathrm{sim}_{\mathrm{pers}}$, the same similarity that was applied to build the neighborhoods was used.

The demographic similarity function $\mathrm{sim}_{\mathrm{dem}}$ was the same for all experiments. It takes into account the country of residence of the users, their age, and their gender and is defined as follows:

$$\mathrm{sim}_{\mathrm{dem}}(a,k) = \frac{1}{\hat{c}+\hat{g}+\hat{a}}$$
$$\left[\hat{c} \cdot \delta_{\mathrm{c}(a)\mathrm{c}(k)} + \hat{g} \cdot \delta_{\mathrm{g}(a)\mathrm{g}(k)} + \hat{a} \cdot \frac{1}{|\mathrm{a}(a) - \mathrm{a}(k)| + 1}\right] \quad (7)$$

In Equation 7, the following helper functions were used: $\mathrm{c}(a)$: returns the country of user a; $\mathrm{g}(a)$: returns the gender of user a; and $\mathrm{a}(a)$: returns the age of user a.

Additionally, the Kronecker delta δ is used. It returns 1 if the two values are equal (e.g., $\mathrm{c}(a) = \mathrm{c}(k)$) or 0 otherwise. The variables $\hat{c}$, $\hat{g}$, and $\hat{a}$ are the weights we give to each of the attributes. The assumption behind this was that in some cases, it might be important to give more importance to one trait than another, especially if, for a given domain, there exists knowledge about the strength of correlations between certain demographic attributes and the type of items that are used in the recommender system. Finally, however, it was decided to use the same weight for every demographic attribute in our experiments (i.e., $\hat{c} = \hat{g} = \hat{a} = 1$).

3.3.2 Personality Factor

The second experiment used the best neighborhood size and similarity measure that was found in the first experiment. Note that these values might vary depending on the approach that is being tested (rating-based vs. personality-based). The goal of this experiment was to determine the best personality factor in the personality-based approach. The RMSE and F1-score were computed for several personality factors p, each time applying 3-fold cross-validation for each user in the dataset. First, it was attempted to look at personality factors with $p = 0, 1, \ldots, 10$. However, it was noticed that this magnitude of change did not reflect itself much in the evaluated metrics. Therefore, it was decided to use larger personality factors with $p = 0, 5, 10, \ldots, 50$. It is also important to note that during this experiment, the demographic factor d was set to 1.

3.3.3 Demographic Factor

The last experiment used different values for the demographic factor d to see whether demographic data could improve prediction accuracy and recommendation relevance. The best configurations found in the two previous experiments were applied, and only the value for d was changed. More precisely, the demographic factor was chosen to be $d = 0, 1, \ldots, 10$. The results of the three experiments are presented in detail in section 4.

3.3.4 Interviews

Besides doing a quantitative analysis of the personality-based collaborative filtering method, opinions from the users of MovieOcean were gathered. A small questionnaire was designed, and interviews were conducted with three of the registered users. Some of the questions asked included: How did you find the questionnaire?, How many questions do you find appropriate for such a questionnaire in this setting?, Did you feel that it captured your personality well?, and How accurate were your recommendations?.

4 RESULTS

This section presents the results of implementing a prototype following the methods explained in the previous section.

4.1 Prototype Implementation

A prototype of MovieOcean was implemented to explore the incorporation of personality traits into a recommender system. This artifact made it possible to collect data from the users and test our personality-based approach in movie recommendations. In the following section, we will give a brief overview of the implementation and features of MovieOcean. The code of MovieOcean, including its evaluation, is available in a public repository on GitHub[3].

In total, 42 users registered on the platform for one month. To complete the registration process, the users needed to answer a few questions about their demographic background. These questions included the age, the country where they lived, and the gender of the user. Most of the users were aged between 20 to 30 years old and lived in Switzerland or Mexico. The majority of the users were women, followed by men, and three users picked another gender variant. After the registration, the personality questionnaire was presented to them. For the users that completed the questionnaire, a personality profile that contains the scores they reached in each personality dimension was provided. Figure 1 presents an example of the personality radar chart shown to the users. Moreover, by issuing a few movie ratings, the users had the chance to improve their recommendations.

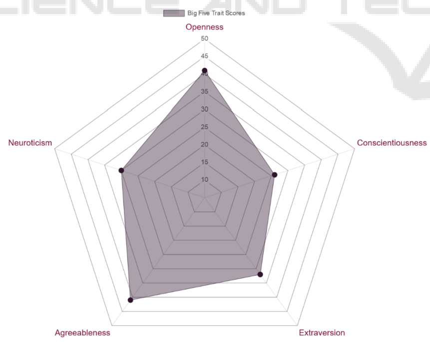

Figure 1: Screenshot of the personality scores that a user would see after completing the questionnaire.

Each movie can be rated on a 5-star scale, and it is possible to give half stars. The minimum rating that can be issued is 0.5 stars. In total, 35 out of the 42 users completed the questionnaire and had, therefore, a personality profile attached to their account. However, only 28 users issued ratings. It was possible to collect a total of 445 movie ratings during the studied period.

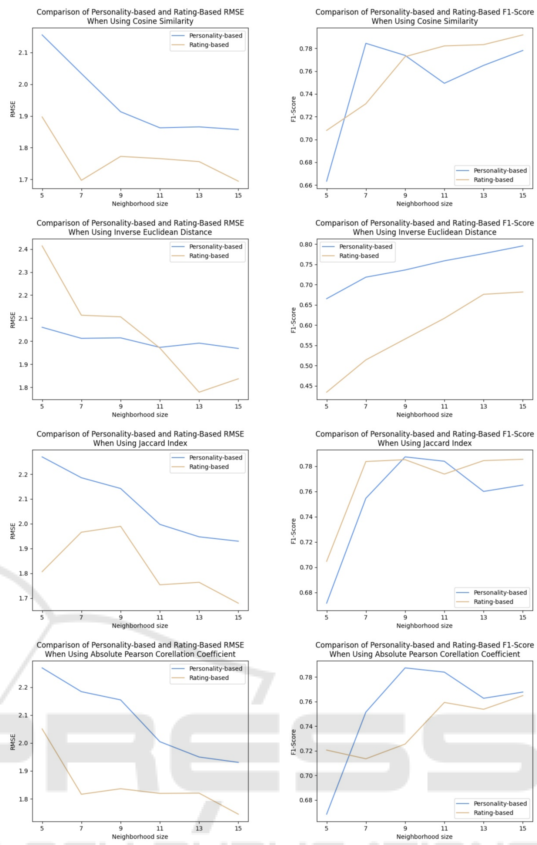

Figure 2: Comparison of the personality-based and rating-based approach when using different similarity measures and different neighborhood sizes.

4.2 Evaluation

The results of the first experiment are summarized in the plots in Figure 2. It can be observed that the metrics improved with a more significant number of neighbors that were considered in the collaborative filtering algorithm. The rating-based approach with cosine similarity achieved the best RMSE of 1.694. The best F1-score was achieved using the personality-based strategy and the inverse Euclidean distance as the similarity measure. The best configuration was unambiguous for the rating-based approach: cosine similarity and a neighborhood size of 15 yielded the best results, both for the RMSE and the F1-score. However, the results were more ambiguous in the personality-based approach: The best RMSE score was achieved using cosine similarity and 15 neighbors. The best F1-score occurred when using the inverse Euclidean distance and 15 neighbors. We decided to use the inverse

[3] available at https://github.com/rolshoven/MovieOcean/

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

Euclidean distance given that, in our opinion, the F1-score is more relevant to the user than the accuracy of the predicted ratings.

Regarding the second experiment, against our initial assumptions, the personality factor did not influence the F1-score of the personality-based approach. However, it did improve the RMSE values (as shown in Figure 3). It could be that an increasing personality factor improves the accuracy for low or high ratings but not much for ratings near the threshold, which would result in better RMSE values but the same F1-scores. On the left-hand side of Figure 3, it can be observed that the RMSE improvement of a more significant personality factor decays exponentially. While increasing the personality factor a lot does not influence the computational complexity, we suggest not choosing a too high value because it will limit the importance of the rating similarity and demographic similarity. A good value for p seems to be somewhere between 30 and 50. In our case, $p = 50$ was chosen before starting the third experiment.

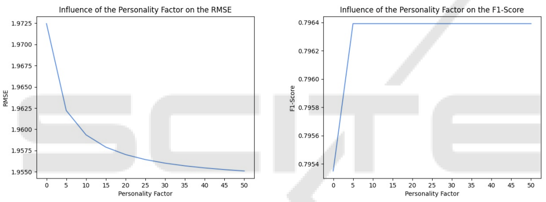

Figure 3: Comparison of the RMSE and F1-Scores using different personality factors.

The last experiment was about the importance of demographics. The results showed that considering additional demographic data during the prediction of movie ratings did not influence the performance of the personality-based approach. We infer that a correlation between personality scores and our demographic attributes (i.e., country, gender, and age) could explain these results. The personality scores could then be seen as a proxy to the underlying demographic data, therefore already considering certain aspects of the demographics implicitly. On the other hand, Figure 4 shows that the demographics influenced the results of the rating-based approach. When we did not consider demographic data, the RMSE value for the rating-based approach was 1.841, whereas it was 1.694 when we set $d = 1$, which is an improvement of 0.147 or almost 8%. Our results suggest that demographic data should be included in a rating-based approach. However, the algorithm should not give too much weight to demographic similarities because it will not further improve the performance. If the recommender system already uses information about the user's personality, there seems to be no need to include additional demographic information.

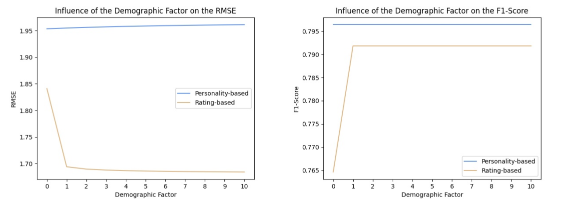

Figure 4: Comparison of the RMSE and F1-scores using different demographic factors.

As a final comparison, a t-test on the RMSE and F1-scores on the rating-based and personality-based approach data in the third experiment with a significance level of $\alpha = 0.05$ was performed. It was done in two scenarios: First, considering no additional demographic data ($d = 0$) and second, using additional demographic data ($d = 1$). The results of this analysis are illustrated in Table 1. There was no significant difference between the F1-scores of the rating-based and the personality-based method. Unfortunately, there was a significant difference between the RMSE values of the two approaches, with the personality-based method performing poorer. The RMSE values for the rating-based approach improved when using additional demographic data, leading to an even smaller p-value.

On the other hand, the personality-based approach does have some advantages over the rating-based approach. Suppose we know that the F1-scores of these two approaches do not differ significantly. In that case, we could assume that the quality of recommendations will not worsen when using the personality-based method.

Table 1: RMSE and F1-scores for the rating-based and the personality-based approach with and without additional demographic data. The last row shows the p-values of the t-test that we performed.

	Without demographics ($d = 0$)		With demographics ($d = 1$)	
Approach	**RMSE**	**F1-Score**	**RMSE**	**F1-Score**
Rating-based	1.874	0.770	1.717	0.799
Personality-based	2.022	0.784	1.955	0.796
p-value	**0.035**	**0.397**	**0.013**	**0.920**

4.3 Interviews Outcome

The conception of recommending movies based on the personality and the implementation of MovieOcean was well-received by users. The design

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

of MovieOcean was perceived as intuitive and aesthetically pleasing. Only one out of the three users had a problem finding a movie but also stated that it could have been because of a typo in the movie name. The users found it easy to navigate through their recommendations and find other movies that were linked as *similar movies* at the end of each movie detail view. Moreover, the personality assessment seemed to reflect what the user thought about themselves. It was mentioned that the overview of the personality scores and the short explanations were received well. Another point is that the star rating system was a good idea and brought some life into the recommender system.

5 SUMMARY AND LESSONS LEARNED

This work explored the inclusion of personality characteristics as an enhancement of user-user collaborative filtering algorithms. To this end, MovieOcean, a personality-based movie recommender system, was implemented. MovieOcean uses a questionnaire based on the Big Five model to generate personality profiles of the users registered on the platform. With the personality profiles, personality-based neighborhoods were created to predict movie ratings and provide recommendations. Furthermore, to assess how the consideration of personality traits influences the recommendations, the RMSE metric was calculated to analyze the accuracy of the predicted ratings and the F1-score to evaluate the relevance of the recommendations for the personality-based and a standard rating-based approach.

Unfortunately, it was not possible to find evidence that incorporating personality in collaborative filtering, in the way it was tested, leads to better performance. The performance is significantly worse in terms of the RMSE of the predicted ratings. However, the F1-scores do not differ significantly, which makes us think that the recommendation quality is the same in both approaches. If this is the case, it is still possible to suggest using the personality-based approach because it has several advantages as per our findings:

1. The neighborhoods of personalities can be precomputed, which leads to a faster recommendation process.
2. It is possible to use correlations between personalities and movie genre tastes to overcome the cold-start problem.
3. Users seem to enjoy the personality-based approach, as the literature and interviews conducted for this study showed.

Regarding the research questions defined in section 1, we conclude that recommender systems cannot rely *more* on a personality-based neighborhood than on a traditional rating-based neighborhood. However, given the case study results, one could claim that it is valid to use personality-based neighborhoods without deteriorating the quality of the recommendations. Moreover, only one method to incorporate personality into recommender systems was explored.

The second question was about the accuracy of the recommendations. Based on our findings, the recommendation quality – measured by the F1-score – does not differ significantly. The significantly worse RMSE of the personality-based approach could also influence the recommendations, but this did not seem to be the case in the conducted experiments. A slightly better F1-Score was found for the personality-based approach, when no additional demographic data was considered. However, when taking demographic data into account, the rating-based method had a slightly higher F1-score. Table 1 shows that in both cases, the difference was not significant. Suppose the F1-score is more important than the RMSE of the predicted ratings. In that case, the authors of this work believe that it is possible to use a personality-based approach without worrying about a negative effect on the recommendation quality. However, based on the obtained results, there are no improvements either.

In addition, the authors consider that the main advantage of personality-based recommender systems is that the neighborhood of similar users can be precomputed, which results in a faster recommendation process. In general, personalities seem to be much more stable than rating vectors, and we do not think the users would frequently change the answers to their questionnaire. Even if they change their answers, they probably only do so a few times, resulting in a low number of recomputations compared to the rating neighborhoods, which might change very often. Moreover, the personality information can be leveraged to solve the cold-start recommendation problem by using correlations between the user personalities and a prototype personality of users that like movies of a specific genre. While many users seem to like the personality-based approach, there are probably also other users that do not want to fill out the entire questionnaire before using a recommender system.

In future work, the impact of different kinds of

Big Five questionnaires on the performance of a personality-based recommender system will be studied. Furthermore, other ways of including personality aspects are going to be explored. For instance, one approach would be to build the neighborhoods using a standard rating similarity and then only consider the users' personality in the similarity measure during the prediction of ratings.

REFERENCES

Aaker, J. L. (1999). The malleable self: The role of self-expression in persuasion. *Journal of marketing research*, 36(1):45–57.

American Psychological Association (2020). Personality. https://www.apa.org/topics/personality. Accessed: 2020-09-19.

Barrick, M. R. and Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, 44(1):1–26.

Braunhofer, M., Elahi, M., and Ricci, F. (2015). User personality and the new user problem in a context-aware point of interest recommender system. In *Information and communication technologies in tourism 2015*, pages 537–549. Springer.

Cantador, I., Fernández-Tobías, I., and Bellogín, A. (2013). Relating personality types with user preferences in multiple entertainment domains. In *CEUR workshop proceedings*. Shlomo Berkovsky.

Choi, K. and Suh, Y. (2013). A new similarity function for selecting neighbors for each target item in collaborative filtering. *Knowledge-Based Systems*, 37:146–153.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., and Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, 40(1):84–96.

Hu, R. and Pu, P. (2013). Exploring relations between personality and user rating behaviors. In *UMAP Workshops*.

John, O. P., Naumann, L. P., and Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy. *Handbook of personality: Theory and research*, 3(2):114–158.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37.

McLaughlin, M. R. and Herlocker, J. L. (2004). A collaborative filtering algorithm and evaluation metric that accurately model the user experience. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 329–336.

Mulyanegara, R. C., Tsarenko, Y., and Anderson, A. (2009). The big five and brand personality: Investigating the impact of consumer personality on preferences towards particular brand personality. *Journal of brand management*, 16(4):234–247.

Nalmpantis, O. and Tjortjis, C. (2017). The 50/50 recommender: a method incorporating personality into movie recommender systems. In *International Conference on Engineering Applications of Neural Networks*, pages 498–507. Springer.

Paiva, F. A., Costa, J. A., and Silva, C. R. (2017). A personality-based recommender system for semantic searches in vehicles sales portals. In *International Conference on Hybrid Artificial Intelligence Systems*, pages 600–612. Springer.

Pervin, L. A. and John, O. P. (1999). *Handbook of personality: Theory and research*, chapter 4, pages 102–105. Guilford Press, 2 edition.

Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41(1):203–212.

Rentfrow, P. J. and Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. *Journal of personality and social psychology*, 84(6):1236.

Tkalčič, M., Kunaver, M., Tasič, J., and Košir, A. (2009). Personality based user similarity measure for a collaborative recommender system. In *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real world challenges*, pages 30–37.

Weaver III, J. B. (1991). Exploring the links between personality and media preferences. *Personality and individual differences*, 12(12):1293–1299.

Zhong, B., Hardin, M., and Sun, T. (2011). Less effortful thinking leads to more social networking? the associations between the use of social network sites and personality traits. *Computers in Human Behavior*, 27(3):1265–1271.