An Approach to a Group Movie Recommender System using Matrix Factorization-based Collaborative Filtering

Faiha Adzra Darmawan[1] and Z. K. A. Baizal[2,*]

[1,2]School of Computing, Telkom University, Bandung 40257, Indonesia

*Corresponding email: baizal@telkomuniversity.ac.id

Abstract: The growth of online movie streaming platforms has driven the demand for recommender systems that can deal with the daunting challenge of users finding movies that match their preferences. However, these recommender systems tend to focus on the needs of individual users, whereas, in the real world, there are circumstances in which recommendations are needed for a group of users. Therefore, this study proposes a group recommender system (GRS) using matrix factorization (MF) with an aggregation model to recommend movies for a group of users. We employ three Matrix Factorization methods, *i.e.*, after factorization (AF), before factorization (BF), and weighted before factorization (WBF) for three distinct group sizes: small, medium, and large. Our goal is to identify the most accurate method for each group size. To evaluate the performance, we used precision and recall as measurement metrics. Based on the evaluation results, the AF method is to outperform the BF and WBF methods in terms of accuracy, both for small and large groups. This can be seen from the precision and recall values which tend to be higher. Meanwhile, the BF method outperforms the AF and WBF methods in terms of precision for medium groups.

Keywords: group recommender system, matrix factorization, movie recommender, user preferences

1 Introduction

In recent years, the evolution of online platforms for movie streaming has led to an abundance of available content, posing a challenge for users to discover movies that align with their preferences. Recommender systems have emerged as a solution to enhance the user

experience by providing personalized suggestions. In response to the increasing demand for recommendations within group settings, the exploration of Group Recommender Systems (GRS) has become a prominent area of research [1–4]. These systems aim to identify items of interest for the entire group, ranging from restaurants and TV shows to music and movies. According to Zhang *et al.* [5], effectively tailored group recommendations can significantly enhance societal interactions. However, suggesting items that cater to the diverse needs of group members remains a challenge, particularly when the group comprises individuals with varying preference [6].

As movie streaming services become increasingly prevalent, there is a growing need for advanced recommender systems. Conventional collaborative filtering techniques predominantly center on individual preferences, potentially overlooking the nuances of group dynamics and shared interests. Acknowledging this shortfall, we aim to make a meaningful contribution by designing a recommender system centered around group dynamics, addressing the changing requirements of users participating in communal movie-watching activities. Matrix factorization (MF) emerges as a formidable technique in the domain of recommender systems. This study introduces a groundbreaking approach to a group movie recommender system utilizing MF-based collaborative filtering, aiming to overcome the limitations inherent in recommendation methods primarily focused on individual preferences. Notably, our study adopts the innovative strategy of incorporating aggregation models, wherein recommendations are not aggregated for individual users [7]. Instead, the approach constructs a comprehensive group preference model, often referred to as a group profile. This group profile is subsequently utilized to determine recommendations, offering a valuable framework in scenarios where group members need to collaboratively analyze, negotiate, and adapt their collective preferences [8].

Group recommender systems have been widely implemented across various domains, including travel suggestions [9, 10], news [11], music [12], books [13], [14], tourism [15], and more. One of the domains that GRS has been implemented in is the offline retail business [1]. This study aims to enhance a GRS by considering the number of recommendations and repeat purchases. It evaluates model performance using various metrics and proposes new evaluation measures, specifically targeting the novelty of the recommender. Another study venture focuses on developing a GRS for selecting tourist destinations tailored to the needs of each group member [3]. This system utilizes a Hybrid Filtering method that combines Collaborative Filtering with Knowledge-Based Filtering. In the academic realm, a study was dedicated to creating a recommender system to gather a panel of experts to evaluate a particular academic pursuit using Collaborative Filtering and Knowledge-Based Filtering [4].

Although many studies utilize MF methods in the scope of recommender systems, these studies focus on individual recommendations. There are many circumstances in which recommendations need to be given to a group of people. Therefore, we built a group recommender system by utilizing three MF methods (AF, BF, WBF) in GRS, to be applied to three different group sizes. This approach aims to address the limitations of individual-centric recommender approaches, introducing a novel dimension to the evolving landscape of recommender systems.

In this research, our primary aim is to introduce a group movie recommender system employing MF-based collaborative filtering integrated with an aggregation model. The focus of our study is the evaluation and comparison of the accuracy of our modified method in delivering pertinent movie recommendations for groups of users. By emphasizing these

objectives, we intend to make a meaningful contribution to the dynamic field of recommender systems, enhancing the personalized and enjoyable group movie-watching experiences for users.

2 Research Method

2.1 Dataset

We utilized the *MovieLens-100K* dataset obtained from the GroupLens website. This dataset comprises 100,000 ratings from 943 users on 1,682 films, with each user providing ratings for at least 20 films. Ratings are on a scale from 1 to 5, where higher values indicate a more positive evaluation. The dataset is tab-separated and lacks column names, necessitating preprocessing for the assignment of column names. We read and preprocess the dataset, providing column names such as *'user_id'*, *'item_id'*, *'rating'*, and *'timestamp'*.

2.2 Group Parameters

Three MF methods, *i.e.*, after factorization (AF), before factorization (BF), and weighted before factorization (WBF) applied in three different scenarios: when the group size is small, medium, and large. These groups are formed randomly from among the users in the *MovieLens-100K* dataset.

(a) Small group: Small groups consist of a limited number of members, which is 2 to 4 users.

(b) Medium group: Medium groups consist of a moderate number of members, which is 5 to 8 users.

(c) Large group: Large groups comprise a substantial number of members, which is 9 to 12 users.

2.3 Creating Group

In this study, we use a dataset that does not contain group preferences yet; hence, we model multiple (randomly generated) groups to generate recommendations that are more relevant to the collective characteristics and preferences of the group members. In the process of generating a group, it is essential to ensure that the items are substantial for effective testing, yielding optimal evaluation outcomes for each approach in this study. We set a testing threshold, referred to as a 'testable threshold' at 50, indicating that at least 50 movies in the test dataset must be rated by at least one member of the group. These groups are non-disjoint, allowing the same user to be part of multiple groups. Subsequently, these groups are classified into three categories: small groups (2-4 users), medium groups (5-8 users), and large groups (9-12 users). Throughout the group formation, our goal is to guarantee that the items possessed are sufficient for comprehensive testing, thereby obtaining optimal evaluation results for each approach in this study.

2.4   Collaborative Filtering

The collaborative filtering (CF) approach represents a method within recommender systems designed to predict the utility of items based on the viewing history of previous users. These algorithms leverage patterns in rating behaviors across multiple users to determine how to suggest items [16]. As outlined in a study [17], CF can be broadly categorized into two primary approaches. There are two main formulas utilized in the CF algorithm [18], as shown on (1) and (2).

(a) Similarity

$$\text{sim}(a, b) = \frac{\sum_{s \in S_{ab}} r_{a,s} r_{b,s}}{\sqrt{\sum_{s \in S_{ab}} r_{a,s}^2} \sqrt{\sum_{s \in S_{ab}} r_{b,s}^2}} \tag{1}$$

Where, $\text{sim}(a, b)$ indicates the similarity value between user a and user b, while $r_{a,s}$ and $r_{b,s}$ represent the rating values from user data with similar preferences.

(b) Rating Prediction

$$r_{c,s} = k \sum_{c' \in C} \text{sim}(c, c') \times r_{c',s} \tag{2}$$

In this context, $r_{c,s}$ signifies the predicted value for the data to be rated, k is the resultant similarity value, $\text{sim}(c, c')$ represents the similarity value between two users and $r_{c',s}$ denotes the rating value in user data with similar preferences.

2.5   Matrix Factorization

Matrix factorization (MF) serves as a method employed in collaborative filtering to address issues of scalability and data sparsity [2, 19]. It is utilized to fill in the gaps or sparsely populated entries within a rating matrix. By estimating the missing values, personalized and accurate recommendations can be generated based on the similarity between users and items in latent factor space [2].

The procedure involves the training of the MF model, considering the training rating $r_{u,i}$ assigned by user u to item i. This rating serves as a pivotal element, with μ denoting the average rating of the dataset, and λ acting as a parameter regulating the training procedure. After mastering the MF, predictions are generated for both individual users and groups, relying on factor vectors, biases, and other model parameters. Precisely, for the prediction of user u to item i, a learned expression is applied. In the case of groups, additional factors and biases are computed, facilitating predictions for the entire group about a specific item [20]. Recommendations for the group are subsequently derived from the pool of unrated items, ensuring the alignment of predictions with the acquired insights of the model. As shown in Figure 1, this system takes the group size as input, generates random groups, and provides a movie recommender based on the group members' biases. The different sizes of groups being considered are small (2-4 users), medium (5-8 users), and large (9-12 users). We use the following methods to implement this system.

(a) AF (after factorization): Figure 2 shows AF method on GRS. In the AF, the system initiates the factorization of the item-by-user matrix and calculates factors for each
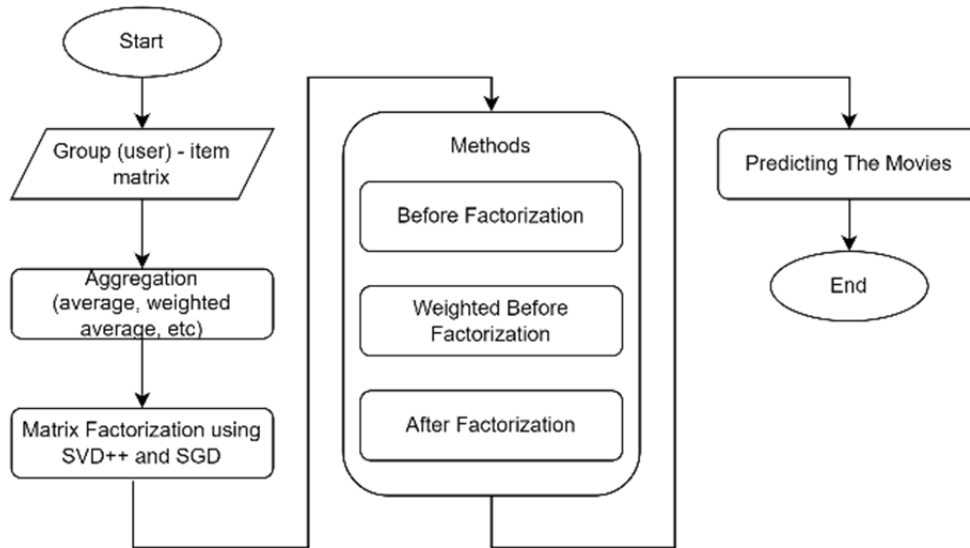
Figure 1: Schematic of GRS movie using MF.

user. Subsequently, it aggregates these factors to determine group factors for the user group. Group rankings are then computed using these group factors for each item, relying on factors and biases obtained from the factorization process.
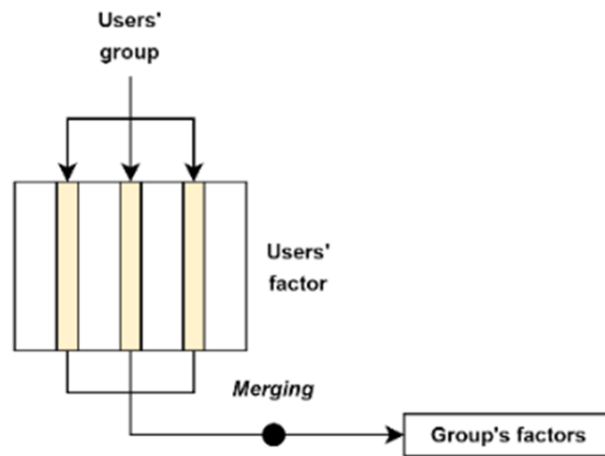

Figure 2: Schematic of GRS movie using AF.

(b) BF (before factorization): In the BF method, we collect user rankings before initiating the factorization on the ranking matrix as shown in Figure 3. We calculate group rankings by leveraging individual rankings from users within the group, treating the

group as a virtual user in this process. Subsequently, we employ ridge regression to obtain factors and biases for the group, encapsulating the fundamental concept of this approach.
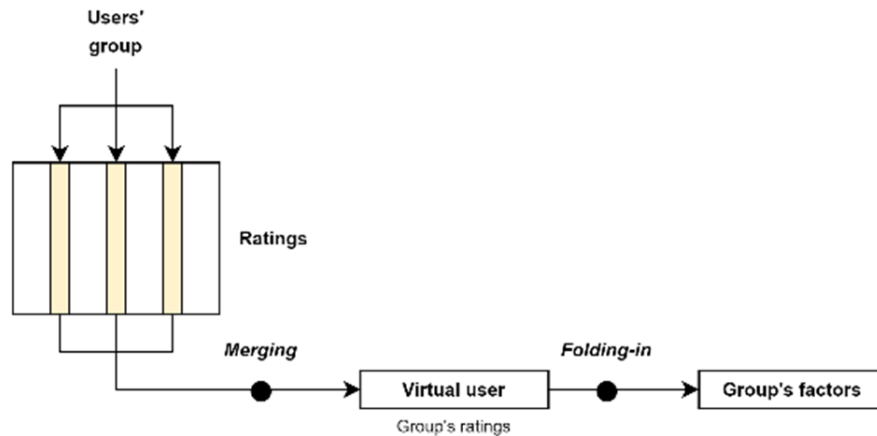

Figure 3: GRS approach with MF method; BF.

(c) WBF (weighted BF): This method resembles the BF method but introduces weighting in the calculations. Figure 4 shows that each calculation assigns a weight to every item based on the number of users who have rated that item and the alignment of their ratings. These weights result in a variation of the gradient minimization function used, considering the weights in each calculation during optimization steps.
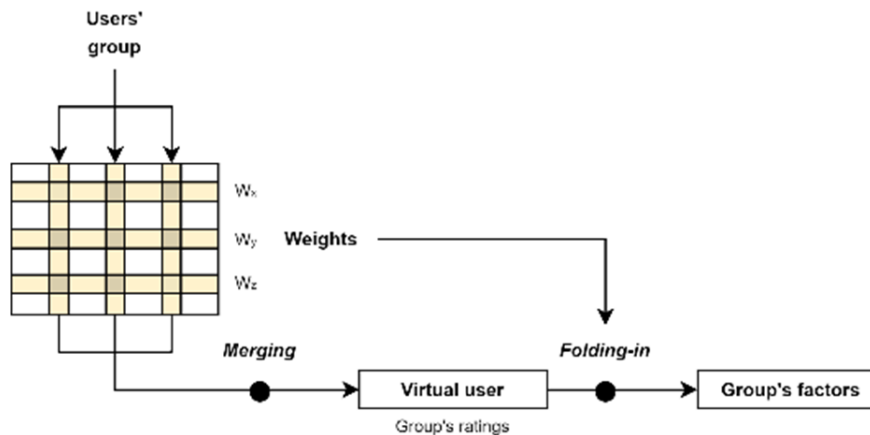

Figure 4: GRS approach with MF method; WBF.

3   Result

In this section, we delve into the results of the experiment. First, we explore the outcomes of implementing the utilized approach. After providing recommendations, finally, we analyze the performance of the models by evaluating the system using precision and recall parameters for AF, BF, and WBF methods across various group sizes.

3.1   Proposed Approach Implementation

In MF, we implement SVD++ (Singular Value Decomposition) and SGD (Stochastic Gradient Descent) methods to decompose the user-item matrix into the user latency matrix and item latency matrix. The demonstration comprises only three iterations, as indicated in Table 1, resulting in a higher mean squared error (MSE) value. However, we conducted additional iterations to achieve a lower MSE for the results.

Table 1: MSE value for each iteration

Iteration	Training MSE	Test MSE
0	0.810	1.045
1	0.755	0.990
2	0.726	0.973

Overall, the decreasing trend in both training and test MSE values across iterations is a positive sign, indicating that the model is learning and making better predictions with each iteration.

3.2   Evaluation Metrics

After providing recommendations, we evaluate to ensure the quality and performance of the system. We assess the three utilized methods, *i.e.*, AF, BF, and WBF, using precision and recall parameters for each group size. To scrutinize the computed recommendation quality for user groups, we employ precision and recall for the defined group G as outlined in (3) and (4).

$$\text{precision}_G = \frac{TP_G}{(TP_G \bigcup FP_G)} \tag{3}$$

$$\text{recall}_G = \frac{TP_G}{T_G} \tag{4}$$

Here, TP_G, FP_G, and T_G represent the sets of true positive, false positive, and expected recommendation sets, as shown in (5), (6), and (7), respectively.

$$FP_G = \{i \in R_G | \exists g \in G \quad \text{where} \quad \hat{r}_{g.i} < 0\} \tag{5}$$

$$TP_G = \{i \in R_G | \exists g \in G \quad \text{where} \quad \hat{r}_{g.i} \neq \bullet \quad \text{and} \quad \forall u \in G \hat{r}_{u.i} \neq \bullet \hat{r}_{u.i} \rightarrow \geq 0\} \tag{6}$$

$$T_G = \{i \in I | \exists g \in G \quad \text{where} \quad \hat{r}_{g.i} \neq \bullet \quad \text{and} \quad \forall u \in G \hat{r}_{u.i} \neq \bullet \hat{r}_{u.i} \rightarrow \geq 0\} \tag{7}$$

Here, R_G signifies the collection of items suggested to group G, while $\hat{r}_{u.i}$ refers to the test rating assigned by user u for item i. These test ratings are not employed in the recommendation computation stage. The parameter θ is utilized as a threshold to ascertain whether a user expresses approval or disapproval of the item.

Simply put, we classify movie recommendations as positive if the test rating surpasses the user satisfaction threshold of 4 for all users in the group, with at least one user having a specified test rating (the test dataset is notably sparse). To assess performance, we implement all three methods (AF, BF, and WBF) on 50 randomly generated non-disjoint groups from users within the *Movielens-100K* dataset.

3.3   Evaluation Results

After evaluating the group movie recommender system, we obtained the evaluation results in terms of precision and recall parameters shown in Table 2 and Table 3, respectively.

Table 2: Precision parameter evaluation results of 3 methods for each group

MF Method	Small (2-4 users)	Medium (5-8 users)	Large (9-12 users)
AF Method	0.856	0.778	0.806
BF Method	0.551	0.805	0.741
WBF Method	0.630	0.645	0.754

Table 3: Recall parameter evaluation results of 3 methods for each group

MF Method	Small (2-4 users)	Medium (5-8 users)	Large (9-12 users)
AF Method	0.074	0.068	0.044
BF Method	0.027	0.026	0.036
WBF Method	0.033	0.025	0.038

Based on the evaluation results in Table 2 and Table 3, it is evident that the AF method demonstrates the best performance for small groups comprising 2-4 users, followed by the WBF method. Here, the SGD method allows AF to iteratively optimize the factorized matrices, this can be attributed to the effective utilization of SVD, particularly AF, where the factorization process enhances the system's ability to capture latent features and patterns within user-item interactions. Similarly, for large groups that consist of 9–12 users, AF continues to outperform both BF and WBF in providing recommendations.

In medium groups (5-8 users), the BF method performs superior accuracy and surpasses both AF and WBF. SVD applied to BF, then the SGD function used in BF contributes to its accuracy in scenarios involving 5-8 users. Increasing the size of medium groups should potentially enhance the success margin of BF even further.

The WBF method, in which each item is associated with weights that depend on the number of users who have watched it and how similar their ratings are, introduces a distinct SGD minimization function. Based on the evaluation results, WBF performs worse than AF and BF for all three group sizes.

4   Discussion

Overall, AF demonstrates superior performance in precision parameters, achieving the highest score of 0.86 in small groups (2-4 users). Subsequently, the BF method surpasses both AF and WBF in precision parameters for medium groups (5-8 users) with a score of 0.81 and reports better accuracy, which is shown in Figure 5.
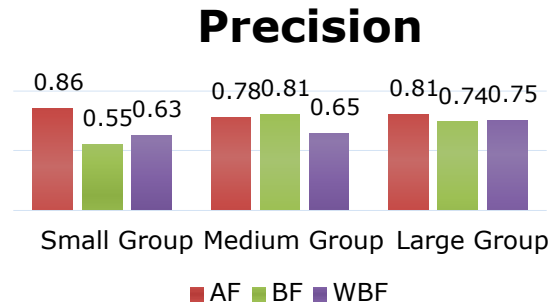

Figure 5: Performance comparison of the precision parameter.

Increasing the group size to a larger scale has the potential to significantly strengthen the success margin of the BF method. The BF method, which is characterized by applying singular value decomposition (SVD) in BF, has shown remarkable accuracy, especially in medium groups. With the anticipation that the success of BF is likely to expand as group size increases, larger group settings offer a broader set of user data, allowing the method to utilize a richer set of latent factors.
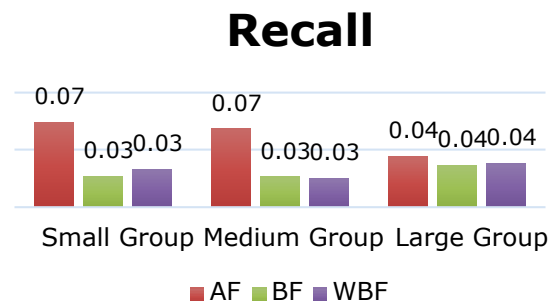

Figure 6: Performance comparison of recall parameter.

We observe that even in terms of recall, which is shown in Figure 6, AF outperforms both BF and WBF methods. Based on the results of the precision and recall values of the model in evaluating the recommender provided by the system to user groups, we note that the WBF method, which incorporates a weighting mechanism, showed lower performance compared to the AF and BF methods across all three group sizes. Nevertheless, it is essential to note that the overall recall values for all three methods remain relatively low. This

could be attributed to the highly sparse dataset, where numerous films in our recommender lack rating data for any group member in the dataset (given that we only report the top 50 recommendations per group in this study). Consequently, we are unable to compute precision or recall for these films, as we cannot determine whether to interpret them as positive or negative matches. In summary, our total set of True Positives is unknown due to the sparsity of data.

The study by Ortega *et al.,* reports much better results for BF and WBF with small and large groups. This could be due to the datasets used are different. The authors used the MovieLens-1MB dataset while we used the *MovieLens-100K* dataset. The author mentions that when the data is small, AF works quite well, as it did on our *Movielens-100 K* dataset. As the data grows, more data is involved in the group recommendation process and the AF method does not work as well. Virtual users are a better representation for user groups than user factor aggregation in this case. Both BF and WBF provide better recommendations for non-sparse datasets.

5 Conclusion

In this study, we built a group recommender system (GRS) using matrix factorization (MF) based collaborative filtering with an aggregation model. We use the MovieLens-100K dataset sourced from the GroupLens website to implement these three MF methods, *i.e.*, AF, BF, and WBF to the three different group categories. These groups consist of varying numbers of members, which are small groups comprising 2-4 users, medium groups consisting of 5-8 users, and large groups accommodating 9-12 users. We compared the methods to determine the most accurate one for each group. Based on the experiments, the AF method is to outperform the BF and WBF methods in terms of accuracy, both for small and large groups. This can be seen from the precision and recall values which tend to be higher. Meanwhile, the BF method outperforms the AF and WBF methods in terms of precision for medium groups. The accuracy of GRS with MF-based CF, particularly in the AF method in smaller groups can be attributed to the model's ability to capture preferences in limited user space, resulting in more accurate and personalized recommendations on a group basis. However, we note that the collective recall values across the three methods persistently exhibit a state of relatively low value. This could be attributed to the highly sparse dataset, where numerous films in our recommender lack rating data for any group member in the dataset (given that we only report the top 50 recommendations per group in this study).

References

[1] J. Park and K. Nam, "Group recommender system for store product placement," *Data Mining and Knowledge Discovery*, vol. 33, no. 1, pp. 204–229, 2019.

[2] F. Ortega, A. Hernando, J. Bobadilla, and J. H. Kang, "Recommending items to group of users using matrix factorization based collaborative filtering," *Information Sciences*, vol. 345, pp. 313–324, 2016.

[3] L. Ambarwati and Z. Baizal, "Group recommender system using hybrid filtering for tourism domain," *Indonesia Journal on Computing (Indo-JC)*, vol. 4, no. 2, pp. 21–30, 2019.

[4] D. T. Hoang, N. T. Nguyen, and D. Hwang, "A group recommender system for selecting experts to review a specific problem," in *Computational Collective Intelligence: 10th International Conference, ICCCI 2018, Bristol, UK, September 5-7, 2018, Proceedings, Part I 10*, pp. 270–280, Springer, 2018.

[5] J. S. Zhang, M. Gartrell, R. Han, Q. Lv, and S. Mishra, "Gevr: An event venue recommendation system for groups of mobile users," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 1, pp. 1–25, 2019.

[6] B. Alhijawi, G. Al-Naymat, N. Obeid, and A. Awajan, "Novel predictive model to improve the accuracy of collaborative filtering recommender systems," *Information Systems*, vol. 96, p. 101670, 2021.

[7] A. Felfernig, L. Boratto, M. Stettinger, and M. Tkalčič, *Group Recommender Systems - An Introduction*. 05 2018.

[8] A. Jameson and B. Smyth, *Recommendation to Groups*, pp. 596–627. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[9] T. D. Pessemier, J. Dhondt, and L. Martens, "Hybrid group recommendations for a travel service," *Multimedia tools and applications*, vol. 76, pp. 2787–2811, 2017.

[10] Z. Xu, X. Zheng, H. Zhang, and Y. Luo, "A novel travel group recommendation model based on user trust and social influence," *Mobile Information Systems*, vol. 2021, pp. 1–10, 2021.

[11] H. Xu, Q. Peng, H. Liu, Y. Sun, and W. Wang, "Group-based personalized news recommendation with long-and short-term fine-grained matching," *ACM Transactions on Information Systems*, vol. 42, no. 1, pp. 1–27, 2023.

[12] A. Valera, Á. Lozano Murciego, and M. N. Moreno-García, "Context-aware music recommender systems for groups: A comparative study," *Information*, vol. 12, no. 12, p. 506, 2021.

[13] E. Ahmed, A. Letta, *et al.*, "Book recommendation using collaborative filtering algorithm," *Applied Computational Intelligence and Soft Computing*, vol. 2023, 2023.

[14] J. Ye, H. Xiong, J. Guo, and X. Meng, "A group recommender system for books based on fine-grained classification of comments," *The Electronic Library*, vol. 41, no. 2/3, pp. 326–346, 2023.

[15] P. Alves, H. Martins, P. Saraiva, J. Carneiro, P. Novais, and G. Marreiros, "Group recommender systems for tourism: how does personality predict preferences for attractions, travel motivations, preferences and concerns?," *User Modeling and User-Adapted Interaction*, vol. 33, no. 5, pp. 1141–1210, 2023.

[16] D. Bokde, S. Girase, and D. Mukhopadhyay, "Matrix factorization model in collaborative filtering algorithms: A survey," *Procedia Computer Science*, vol. 49, pp. 136–146, 2015.

[17] R. R. Konguasa, "rancang bangun aplikasi rekomendasi wisata museum di jakarta dengan metode item based collaborative filtering berbasis android," 2018.

[18] S. H. Ati, R. Saptono, and U. Salamah, "Peningkatan efektivitas metode user-item based collaborative filtering pada sistem rekomendasi wisata kuliner kota solo," *ITS-MART: Jurnal Teknologi dan Informasi*, vol. 1, no. 1, pp. 01–04, 2012.

[19] R. Duan, C. Jiang, and H. K. Jain, "Combining review-based collaborative filtering and matrix factorization: A solution to rating's sparsity problem," *Decision Support Systems*, vol. 156, p. 113748, 2022.

[20] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.