Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.5457

RecSys-Fairness: A Framework for Reducing Group Unfairness in Recommendations

Rafael Vargas Mesquita dos Santos [**Federal Institute of Espírito Santo** | ***rafaelv@ifes.edu.br***]
Giovanni Ventorim Comarela [**Federal University of Espírito Santo** | ***gc@inf.ufes.br***]

Institute of Computing, Federal Institute of Espírito Santo, Cachoeiro de Itapemirim, ES, 29322-000, Brazil.

Received: 20 January 2025 • **Accepted:** 17 September 2025 • **Published:** 21 February 2026

Abstract. In this study, we address the importance of promoting fairness in recommendation systems, which are highly susceptible to biases that can lead to unfair outcomes for different user groups. We developed a fairness algorithm aimed at mitigating these injustices, which was applied to the MovieLens dataset and analyzed based on the recommendations produced by the ALS (Alternating Least Squares) and NCF (Neural Collaborative Filtering) methods. Users were grouped by activity level, gender, and age, and the results demonstrated the effectiveness of the fairness algorithm in substantially reducing group unfairness (R_{grp}) across all tested configurations, without causing significant losses in recommendation accuracy, measured by the Root Mean Squared Error ($RMSE$). In particular, a reduction in group unfairness of up to 65.57% was observed in the ALS method. Additionally, we identified an optimal convergence of the fairness algorithm for an estimated number of matrices (h) between 10 and 15, suggesting an effective balance point between promoting fairness and maintaining precision in recommendations. In comparison with the available benchmarks, under identical experimental conditions, we managed to improve group unfairness reductions by approximately 6% (from 59.77% to 65.57%).

Keywords: Recommender Systems, Fairness, Unfairness, Individual Fairness, Group Fairness

1 Introduction

Recommendation systems are decisive to various online platforms that exert significant influence over the choices we make in our everyday lives. From social media platforms such as Facebook and Twitter to streaming services such as Netflix and transportation apps such as Uber, these systems shape our preferences and decisions. However, as our reliance on these systems increases, it is important to consider potential inadvertent social harms that may arise.

Computational models are not free from bias because their algorithms and training data have specific limitations [Taso *et al*., 2023; Deldjoo *et al*., 2021]. These models may present biases and privilege certain groups over other groups [Ruback *et al*., 2021]. Therefore, due to non-neutrality they can make discriminatory decisions [Niemiec *et al*., 2022].

In recent studies, it has been emphasized how recommendation systems, by predicting user preferences, can unintentionally perpetuate inequalities and unfairness. For instance, Beutel *et al*. [2017] and Burke *et al*. [2018] indicate the possibility of such systems offering unfair or unequal quality of service to certain individuals or user groups. Furthermore, it is important to emphasize that these systems can also contribute to social polarization, widening the divergence between individual or user group preferences, as demonstrated by Dandekar *et al*. [2013].

As our dependence on recommendation systems grows, it is essential that we are aware of potential negative impacts and that we work to mitigate them in order to promote a more fair and inclusive environment.

Recommendation accuracy is often used as a metric to evaluate the performance of a recommendation algorithm-how well it can predict whether a user may like an item or not, i.e., its utility. However, the issue of user fairness arises when it is necessary to consider the unequal effects of recommendations on certain groups.

Collaborative filtering-based recommendation systems rely on user-provided data to learn models that are used to predict users' unknown preferences. However, the recommendations generated by these systems can reproduce undesired characteristics inherent to the observed data.

In this article, we utilize the definitions of fairness measures proposed by Rastegarpanah *et al*. [2019]. Based on these definitions, we implement a fairness strategy aimed at reducing unfairness among user groups, ensuring a more balanced and equitable distribution of recommendations.

Furthermore, we examine the relationship between improvements in socially relevant measures and changes in the overall system accuracy. Our analysis extends to exploring the intricate balance between enhancing fairness and maintaining high levels of accuracy.

This article is organized into four additional sections to provide greater clarity and structure. Section 2 is dedicated to presenting the results of a comprehensive literature review, offering the theoretical foundation necessary for the study. In Section 3, we describe in detail the materials used, the datasets analyzed, the proposed methodological approach, and the experimental procedures adopted. Section 4 presents the experimental results obtained, accompanied by an in-depth discussion of their relevance and limitations. Finally, Section 5 presents the final considerations and suggests possible directions for future research.

2 Related Work

Justice is a topic of growing interest in the field of machine learning, especially in the context of recommendation systems. We consider a recommendation system to be fair if it minimizes disparities in quality metrics, such as accuracy or mean squared error, between different sensitive groups [Kamishima *et al*., 2012; Dwork *et al*., 2011]. Fairness can also be understood as the absence of bias in the decision-making process, ensuring that sensitive attributes such as race, gender, or age do not disproportionately influence outcomes [Zemel *et al*., 2013]. Furthermore, fairness may involve balancing the trade-off between individual fairness, which ensures similar treatment for similar individuals, and group fairness, which seeks equitable treatment across predefined sensitive groups [Hardt *et al*., 2016]. Recent studies have also emphasized the importance of fairness in dynamic environments, where recommendation systems must adapt to evolving user preferences while maintaining equitable treatment [Beutel *et al*., 2017].

However, despite advancements, many studies in the field still present significant limitations. For example, approaches that consider group-level fairness often face difficulties in precisely defining sensitive groups and practically applying balancing criteria [Kamishima *et al*., 2012]. Individual fairness approaches, on the other hand, are challenged by the complexity of ensuring consistency among users with similar characteristics [Dwork *et al*., 2011].

Additionally, studies such as those by Burke *et al*. [2018] highlight the importance of a bidirectional perspective on fairness, considering both the experience of users and the items being recommended. Nevertheless, these approaches often require frequent modifications to the learning model, which can limit their application in large-scale systems [Burke *et al*., 2018].

To ensure robust evaluation, we link fairness to well-known statistical metrics, such as variance and mean squared error, enabling a quantitative analysis of the uniformity in service quality [Beutel *et al*., 2017].

In relation to research on learning tasks such as classification and regression, few researchers have explored notions of fairness in the context of recommendation systems. Recently, Burke *et al*. [2018] observed that recommendation systems predicting users' preferences for items should consider fairness from both sides: the perspective of users receiving recommendations and the perspective of items being recommended. Some of the early works by Kamishima and Akaho [2017]; Kamishima *et al*. [2012, 2018] focused on group-level fairness notions, modifying the learning model to ensure that item recommendations were independent of user characteristics, such as race and gender. More recently, Beutel *et al*. [2017] and Yao and Huang [2017] defined group-level fairness notions in recommendation systems based on prediction accuracy across different user or item groupings.

Finally, it is important to mention that despite significant contributions, much literature does not address strategies for reducing disparities in service quality in real-time systems. Moreover, few approaches explore post-processing solutions to ensure fairness, which limits the practical applicability of many proposals [Rastegarpanah *et al*., 2019].

Our contributions:

1. We explore a different approach from most other works to incorporate fairness notions in recommendation systems.
2. We link the fairness metric to well-known statistical concepts, such as variance and mean squared error, to ensure robust evaluation.
3. Our approach avoids direct modifications to the algorithm for each fairness principle, providing greater flexibility and scalability.
4. We propose a post-processing alternative that effectively reduces group disparities.
5. Under identical experimental settings, our approach surpassed the benchmark established by Rastegarpanah *et al*. [2019], achieving an approximate 6% improvement in reducing group disparities (from 59.77% to 65.57%), reinforcing the effectiveness of our proposal compared to existing solutions.

3 RecSys-Fairness Framework

In this section, we will present an overview of the RecSys-Fairness[1] framework. Modules 1 and 2 of the framework will be detailed.

3.1 Module 1: Fairness Measures

We used module 1 of the algorithm to calculate social fairness measures in the proposed case studies. Following the specifications and discussions from the previous section, we formally define the metrics that specify the objective functions associated with individual fairness and group fairness. It is pertinent to mention that all definitions of fairness measures detailed in this section were proposed in the work of Rastegarpanah *et al*. [2019]. In our work, we implemented these definitions and used them as part of our social fairness strategy in recommendation systems. In section 3.2, both the definitions and the implementation are original.

We start by presenting the system configuration, the notation, and the problem definition. Let us suppose that $X \in \mathbb{R}^{n \times d}$ is a partially observed rating matrix of n users and d items, where the element x_{ij} denotes the rating given by user i to item j. Let Ω be the set of indices of known ratings in X. Furthermore, let Ω^i denote the indices of known item ratings for user i, and let Ω_j denote the indices of known user ratings for item j.

For a matrix $\mathbf{A}$, $P_\Omega(\mathbf{A})$ is a matrix whose elements at $(i, j) \in \Omega$ are a_{ij}, and zeros elsewhere. Similarly, for a vector $\mathbf{a}$, $P_{\Omega_j}(\mathbf{a})$ is a vector whose elements at $i \in \Omega_j$ are the corresponding elements of $\mathbf{a}$, and zeros elsewhere. Throughout the paper, we denote the j-th column of $\mathbf{A}$ by the vector $\mathbf{a}_j$ and the i-th row of $\mathbf{A}$ by the vector $\mathbf{a}^i$.

Given a traditional recommendation system, an estimated matrix $\hat{X} = [\hat{X}_{ij}]_{n \times d}$ is generated. In this recommendation problem, we assume users in a set $\{u_1, u_2..., u_n\}$ and items in a set $\{v_1, v_2..., v_d\}$.

[1] https://github.com/ravarmes/recsys-fairness

3.1.1 Individual Fairness

For each user i, we define ℓ_i as the mean squared error estimated over the known ratings, with j indexing the items rated by user i.

$$\ell_i = \frac{\sum_{j \in \Omega^i} (\hat{x}^i_j - x^i_j)^2}{|\Omega^i|} \tag{1}$$

Next, we define individual unfairness as the variation of user losses:

$$R_{indv}(X, \hat{X}) = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l>k} (\ell_k - \ell_l)^2 \tag{2}$$

To enhance individual fairness, we aim to minimize R_{indv}.

3.1.2 Group Fairness

Let I be the set of all users/items and $G = \{G_1, G_2, ..., G_g\}$ be a partition of users/items into g groups, i.e., $I = \bigcup_{k=1}^{g} G_k$. We define the group loss as the estimate of the mean squared error over all known ratings in group k:

$$L_k = \frac{\sum_{(i,j) \in \Omega_{G_k}} (\hat{X}_{i,j} - X_{i,j})^2}{|\Omega_{G_k}|} \tag{3}$$

For a given partition G, the unfairness of the groups is the variation among all group losses:

$$R_{grp}(X, \hat{X}, G) = \frac{1}{g^2} \sum_{k=1}^{g} \sum_{l>k} (L_k - L_l)^2 \tag{4}$$

Again, to improve group fairness, we minimize R_{grp}.

3.2 Module 2: Fairness Algorithm

Module 2 of the algorithm was used to calculate a recommendation matrix that minimizes group unfairness, i.e., that maximizes fairness for social measures in recommendation systems.

We use the estimated matrix $\hat{X}$ to generate h other estimated matrices $\hat{X}_1$, $\hat{X}_2$..., $\hat{X}_h$. These h estimated matrices are generated with random variations, bounded within $-\ell_i/4$ and $+\ell_i/4$, for each value of $\hat{x}^i$ corresponding to user i.

The new values of cells $\hat{x}^i$ in each of the estimated matrices $\hat{X}_p$ can consider a perturbation strategy based on the variance of rating differences versus recommendations. In this context, we set a maximum variance of 16 $(5-1)^2$, as the largest possible difference between an actual and recommended value is 4. For instance, we can consider an actual rating of 1 for a specific item compared to a recommendation for the same item calculated at a value of 5. Thus, we normalize the random recommendation value by dividing it by four times the individual unfairness ℓ_i.

For each estimated matrix, we calculate n individual losses (ℓ_i), corresponding to the n users. Therefore, for each estimated matrix $\hat{X}_p$, where $\{1 \leq p \leq h\}$, we have a list of n individual losses $\{\ell_1, \ell_2..., \ell_n | \hat{X}_p\}$.

We define the matrix of individual losses $Z = [Z_{ij}]_{n \times h}$ to represent the n individual losses calculated for each of the h estimated matrices $\hat{X}_p$, where $Z_{ij} \in \{\mathbb{R}_+\}$, and $\{1 \leq i \leq n\}$, and $\{1 \leq j \leq h\}$, index users and estimated matrices, respectively.

We define the binary matrix $W = [W_{ij}]_{n \times h}$ to indicate whether individual loss j is considered for a user i in forming the final estimated matrix $\hat{X}_\pi$, where $W_{ij} \in \{0, 1\}$, $\{1 \leq i \leq n\}$, and $\{1 \leq j \leq h\}$ index users and individual losses, respectively. Specifically, if individual loss j is considered for user i, then $W_{ij} = 1$; otherwise, $W_{ij} = 0$.

The optimization algorithm is applied using the Gurobi Optimization, LLC [2024] solver to select n rows $\{v_1, v_2, \ldots, v_m | u_i\}$, generating a single estimated matrix $\hat{X}_\pi$. In this algorithm, the goal is to minimize the group unfairness R_{grp}. The matrix W ensures that each row, representing a user, has exactly one value equal to 1. Each 1 in a row of matrix W indicates from which estimated matrix $\hat{X}_p$ the estimates for user i should be selected. The optimization algorithm seeks to find the best combination that minimizes the value of R_{grp}.

Thus, we formulate the optimization procedure for the fairness-aware recommendation problem as follows:

$$R^{min}_{grp} = \frac{1}{g} \sum_{k=1}^{g} (L_k - \mu)^2 \tag{5}$$

where:

$$\ell_i = \sum_{i=1}^{n} \sum_{j=1}^{h} W_{ij} Zij \tag{6}$$

$$L_k = \frac{1}{|\Omega_{G_k}|} \sum_{i \in G_k} \ell_i \tag{7}$$

For a better understanding of the RecSys-Fairness strategy, consider Figure 1 and the Pseudocode of Algorithms 1, 2, and 3. The 6 steps of Figure 1 can be detailed as follows:

1. Prediction: the matrix X, partially filled, is considered by a traditional recommendation system to make recommendation predictions in $\hat{X}$;
2. Clustering: users are grouped based on some common characteristic;
3. Estimated Matrices: h matrices generated by perturbations in $\hat{X}$;
4. Individual Fairness: compute user losses in each of the h matrices to build Z.
5. Group Fairness: using an optimization algorithm, the binary matrix W is created to achieve the fairest combination of recommendations, minimizing R_{grp};
6. Solution Matrix: structure $\hat{X}_\pi$ from W.

4 Materials and Methods

This section describes the methodology employed in the computational experiments conducted in this study. It is structured to provide a comprehensive understanding of the process, from data acquisition to the final analysis, ensuring transparency and reproducibility of the results.

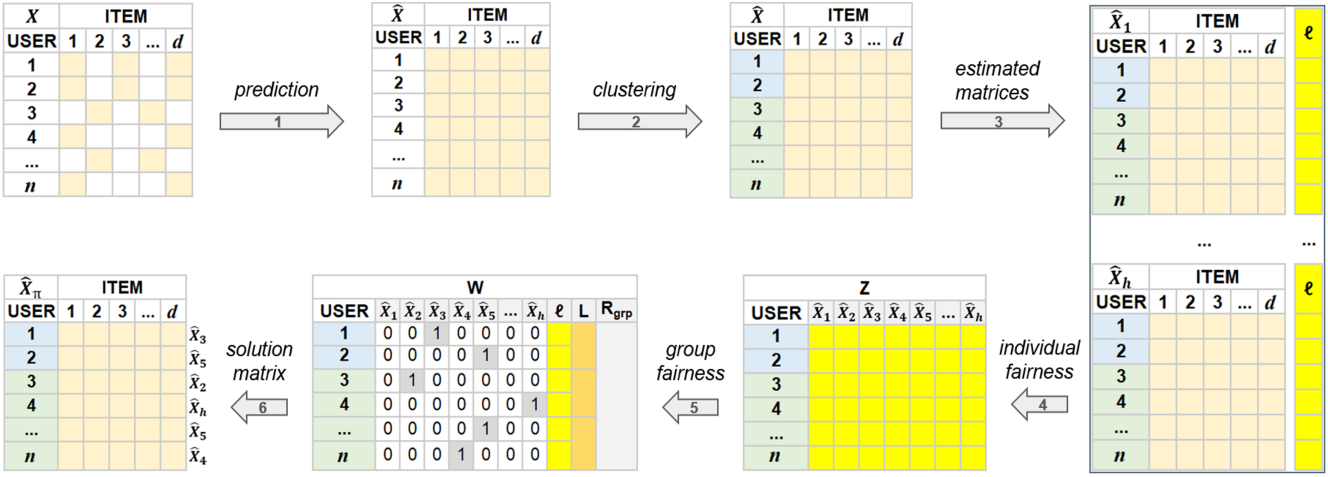

Figure 1. Scheme of the fairness algorithm

Algorithm 1 Fairness Algorithm

Require: Rating matrix $X \in \mathbb{R}^{n \times d}$, groups $G = \{G_1, G_2, \ldots, G_g\}$, matrices count h
Ensure: Estimated matrix $\hat{X}_\pi$ with fairness
Calculate initial $\hat{X}$ using recommendation algorithm
Evaluate R_{indv}, R_{grp} and $RMSE$ of $\hat{X}$
$\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_h\} \leftarrow$ EstimatedMatrices($\hat{X}$, h)
for $p \leftarrow 1$ to h **do**
    Calculate R_{indv}, R_{grp} and $RMSE$ of $\hat{X}_p$
end for
Build loss matrix $Z \in \mathbb{R}^{n \times h}$
$\hat{X}_\pi \leftarrow$ SolutionMatrix(Z, G, $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_h\}$)
return $\hat{X}_\pi$

Algorithm 2 Estimated matrices

Require: Estimated matrix $\hat{X} \in \mathbb{R}^{n \times d}$, matrices count h
Ensure: Set of perturbed matrices $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_h\}$
for $p \leftarrow 1$ to h **do**
    Calculate mean deviations μ per user
    Calculate variances σ and scale by 1/4
    Set lower bound B_{min} as 0 or σ based on μ sign
    Set upper bound B_{max} as σ or 0 based on μ sign
    Adjust bound dimensions for matrix compatibility
    Sample random perturbations from $[B_{min}, B_{max}]$
    Adjust perturbation signs to match original direction
    $\hat{X}_p \leftarrow \hat{X} +$ perturbations
    Clip values in $\hat{X}_p$ to range $[1, 5]$
end for
return $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_h\}$

4.1 Dataset

To test the fairness algorithm, we used the MovieLens 1M dataset[2], which contains approximately 1 million ratings for 3952 movies by 6040 users, with ratings on a 5-point scale [Harper and Konstan, 2015]. We filtered the top 300 users, along with the top 1000 most rated movies. The sparsity of the dataset before and after filtering was 95.81% and 48.03%, respectively.

The filtering of users and items for the experiments was

[2]https://grouplens.org/datasets/movielens/

Algorithm 3 Solution matrix

Require: Loss matrix Z, groups G, matrices $\{\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_h\}$
Ensure: Optimized matrix $\hat{X}_\pi$
Create optimization model
Define binary $x_k[i, j]$ for user i, group k, matrix j
$L_k \leftarrow$ Group-k loss: $\frac{1}{|G_k|} \sum_{i \in G_k} \sum_j Z[i, j] \cdot x_k[i, j]$
$L_{mean} \leftarrow$ Group loss mean: $\frac{1}{g} \sum_{k=1}^{g} L_k$
$R_{grp} \leftarrow$ Group variance: $\frac{1}{g} \sum_{k=1}^{g} (L_k - L_{mean})^2$
Minimize R_{grp}
Add constraint: each user uses exactly one matrix
Solve optimization model
Build $\hat{X}_\pi$ with selected ratings
return $\hat{X}_\pi$

based on the work of Rastegarpanah *et al.* [2019]. We adhered to this configuration to enable a comparison of our algorithm's performance in reducing unfairness with the state-of-the-art benchmarks.

4.2 User Clustering

Users were grouped by activity level (number of ratings), gender, and age, as detailed in Table 1, which shows the user count per cluster group.

- Activity (95-5): one group containing 5% of users with the highest number of ratings, and the remaining 95% of users considered in another group. The 5% group would be the favored users, while the 95% group are the disfavored users;
- Gender: male and female;
- Age: under 18 years old, 18 to 24 years old, 25 to 34 years old, 35 to 44 years old, 45 to 49 years old, 50 to 55 years old, and over 55 years old.

In the proposed framework, each user group is treated with equal weight in the fairness evaluation, regardless of its size. The variance between groups is calculated considering only the mean performance of each group, without weighting by the number of users. This choice was motivated by the goal of promoting a balanced treatment of fairness across all sensitive groups, avoiding the neglect of minority groups.

Table 1. Details of user clustering

Clustering	Groups	Quantity
Activity	Favored	15
	Disfavored	285
Gender	Male	240
	Female	60
Age	Under 18 years old	5
	18 to 24 years old	53
	25 to 34 years old	142
	35 to 44 years old	58
	45 to 49 years old	24
	50 to 55 years old	12
	Over 55 years old	6

However, it is acknowledged that, in certain contexts, weighting the impact of each group by its size could be desirable to reflect the overall representativeness of users in the system. The choice between a weighted or unweighted approach depends on the specific objectives of the application and represents an interesting direction for future investigations.

4.3 Recommendation Algorithms

We also considered two filtering strategies in recommendation systems to estimate unknown ratings. We opted to use ALS (Alternating Least Squares) and NCF (Neural Collaborative Filtering), each with unique approaches that differentiate them from other available methods.

- **ALS (Alternating Least Squares)**: ALS is used to perform matrix factorization in recommendation systems. It decomposes the rating matrix into latent factors, enabling the generation of personalized recommendations. This approach is effective for handling sparse datasets and scales well for large systems. However, determining appropriate parameters can be key, and ALS may face challenges when dealing with very sparse data or new users/items [Hardt, 2013; Hastie *et al*., 2014; He *et al*., 2016].
- **NCF (Neural Collaborative Filtering)**: NCF uses neural networks to model interactions between users and items. This approach allows capturing complex patterns in the data, especially in scenarios where the relationship between users and items is non-linear. NCF is effective in improving the accuracy of personalized recommendations by learning latent representations directly from interaction data [He *et al*., 2017; Wang *et al*., 2018].

These characteristics emphasize why we chose ALS and NCF: ALS is recognized for its effectiveness in traditional recommendation systems by using matrix factorization to generate personalized recommendations based on explicit preference patterns. On the other hand, NCF introduces an innovative approach by employing neural networks to model complex and non-linear interactions between users and items. This advanced methodology aims to significantly enhance recommendation personalization by capturing subtle nuances and patterns in interaction data. These fundamental differences not only reflect technological evolution in the field of recommendation systems but are also significant for understanding how each method may behave uniquely in the practical application of our research.

4.4 Hyperparameter Optimization

The hyperparameter optimization process followed a rigorous methodology to ensure unbiased evaluation. We employed a data partitioning strategy with 80% of the ratings for training and 20% for testing, as described in Section 4.6. Hyperparameter tuning was conducted exclusively on the training set using a 5-fold cross-validation approach, ensuring that the test set remained untouched and was used only for the final model evaluation. This methodology preserved the integrity of the testing process and avoided information leakage.

The optimization process was primarily guided by the Root Mean Square Error (RMSE) as the evaluation metric to ensure prediction accuracy.

For the ALS (Alternating Least Squares) algorithm, we tested {10, 15, 20, 25, 30} values for `rank`, representing the number of latent factors, and {0.1, 1, 10, 20, 50} values for `lambda`, the regularization parameter to prevent overfitting. The optimal configuration was determined to be `rank`=20 and `lambda`=20. The number of iterations was fixed at 15, based on empirical observation that additional iterations did not significantly improve the RMSE on the validation folds. ALS was implemented in the `RecSysALS` class, alternating between user and item matrix updates until convergence.

For the Neural Collaborative Filtering (NCF) model, we evaluated {10, 15, 20, 25, 30} values for `n_factors`, {32, 64, 128} for `batch_size`, and {10, 15, 20, 25} for the number of epochs. The optimal parameters were found to be `n_factors`=20, `batch_size`=64, and `epochs`=20. The model architecture includes embedding layers for users and items, followed by concatenation and dense layers with ReLU activation, and is trained using the Adam optimizer with the mean squared error loss function, implemented in the `RecSysNCF` class.

4.5 Estimated Matrices Count

To determine the optimal number of estimated matrices to be calculated by the fairness algorithm, we explored different values of h: 3, 5, 10, 15, 20 and 25. The aim of varying h is to identify the point at which the algorithm begins to show signs of convergence, which is essential for ensuring the effectiveness and efficiency of the optimization process. The concept of algorithm convergence is fundamental in optimization and machine learning, indicating the point at which further adjustments to the parameters yield marginal or no improvement in the results. Following the guidelines from Bishop [2006] and Goodfellow *et al*. [2016], analyzing convergence is essential for validating the stability and reliability of learning algorithms.

We conducted 10 repetitions for each h value to ensure the robustness of our results. This approach helps mitigate the impact of outliers, which can distort the performance evaluation of the algorithm. Using multiple repetitions and analyzing the average values are recommended practices in simulation and experimental studies to increase the reliability

of the conclusions [Hastie *et al*., 2009; James *et al*., 2013]. The choice of conducting 10 repetitions strikes a balance between the need for statistical precision and computational feasibility, as discussed in recent works by LeCun *et al*. [2015] and Krizhevsky *et al*. [2017].

4.6 Data Splitting for Training and Testing

To evaluate the performance of the recommendation algorithms, we randomly divided the filtered dataset, which consists of 300 users and the 1000 most-rated movies, into two main parts: one for training (80%) and one for testing (20%). This random splitting approach is essential to ensure that the training and testing sets are representative and independent, allowing for a fair and unbiased evaluation of the models. In the training set, the algorithms are adjusted to the available data, enabling them to learn the necessary patterns and structures for making recommendations. The test set, on the other hand, is used to assess the models' ability to generalize to new data, providing an objective measure of their effectiveness in real-world scenarios. This methodology is widely recognized in the machine learning and recommendation literature as fundamental to avoid overfitting and to ensure reliable results [Bishop, 2006; Goodfellow *et al*., 2016].

5 Results and Discussions

In this section, we present the performance of our fairness algorithm, drawing attention to the recommendation quality and the effectiveness of fairness compared to traditional recommendation algorithms that do not consider fairness.

The results of the experiments are presented in tables 2, 3, 4, 5, 6 and 7 providing the following information:

- Algorithm: Name of the algorithm used before the application of the fairness strategy;
- Clustering: Type of user clustering considered;
- h: Quantity of calculated estimated matrices;
- Mean: Mean resulting from 10 repetitions of the fairness algorithm execution;
- Standard Deviation: Standard deviation resulting from 10 repetitions of the fairness algorithm execution;
- (Δ%): Percentage reduction or increase comparing the original value and the mean resulting from the fairness algorithm execution.

The cells with the greatest reductions in group unfairness (R_{grp}) were intentionally highlighted in red. These visual markings emphasize the scenarios showcasing the best outcomes achieved by the proposed fairness algorithm.

The presented tables provide a detailed analysis of the performance of the proposed fairness algorithm, comparing it with traditional recommendation algorithms on the MovieLens 1M dataset, using three distinct user groupings (by activity level, gender, and age) and two traditional recommendation strategies (ALS and NCF). The results are analyzed under two main indicators: group unfairness (R_{grp}), which measures the fairness of the recommendations, and root mean squared error ($RMSE$), which evaluates the effectiveness of the recommendations in terms of accuracy. These configurations were applied for different numbers of estimated matrices (h).

In the **Activity** grouping, we observe that both ALS and NCF algorithms significantly reduce R_{grp} as the number of estimated matrices increases. For example, with ALS, R_{grp} decreases by up to 65.57% at $h = 25$, while NCF achieves a maximum reduction of 37.13% at $h = 15$. Regarding $RMSE$, both algorithms show slight variations, indicating maintained efficiency in recommendations.

In the **Gender** grouping, a similar pattern emerges with significant reductions in R_{grp} as h increases. ALS reduces R_{grp} by up to 40.75% at $h = 25$, while NCF achieves a reduction of 43.43% at $h = 25$. In terms of $RMSE$, both algorithms show slightly higher values than the previous grouping, suggesting a minor loss of performance in customizing recommendations for users grouped by gender after applying the fairness algorithm.

In the **Age** grouping experiments, the results show significant reductions in R_{grp} as the number of estimated matrices increases, using the ALS and NCF algorithms on the MovieLens 1M dataset. With ALS, we observe a progressive reduction in R_{grp} from 0.002170 to 0.001673 at $h = 25$, representing a reduction of 22.91%. Similarly, NCF shows a reduction in R_{grp} from 0.001367 to 0.001003 at $h = 20$, resulting in a reduction of 26.65%. As for $RMSE$, both algorithms show modest increases, very close to the grouping by gender.

Although the absolute values of group unfairness (R_{grp}) are small, it is essential to consider the percentage reduction. This demonstrates the overall capability of the algorithm to reduce unfairness in scenarios with larger absolute values.

In summary, the results demonstrate the proposed fairness algorithm's ability to significantly reduce group unfairness across various algorithm and grouping configurations, with an acceptable compromise in recommendation accuracy. This underscores the importance and effectiveness of integrating fairness considerations into recommendation systems, pointing to promising avenues for future research in the field.

Figure 2 offers a detailed comparison between the metrics of group unfairness reduction (R_{grp}) and the increase in root mean squared error ($RMSE$), considering different clustering strategies (by activity level, gender, and age) and using the ALS and NCF algorithms. Each subplot illustrates the variation of these metrics as a function of the number of estimated matrices (h), allowing us to observe the behavior of the metrics as we adjust the model's complexity. The curves for R_{grp} and $RMSE$ are presented simultaneously for each clustering and algorithm configuration, offering a clear view of how fairness and accuracy evolve with different levels of adjustment in the h parameter. This arrangement facilitates the understanding of the trade-offs involved in the pursuit of fairer recommendation systems without significantly compromising the quality of the recommendations provided.

Observing Figure 2, we delve deeper into the dynamics between the performance of the ALS and NCF algorithms in various clustering strategies, namely, by activity level, gender, and age. The figure contrasts the evolution of group unfairness reduction (R_{grp}) and root mean squared error ($RMSE$) as the parameter h increases. A distinctive pattern emerges,

Table 2. R_{grp} and $RMSE$ Metrics: Activity-Based Clustering with **ALS** on **MovieLens 1M**

Activity-Based Strategy for User Grouping $\{G_1, G_2\}$						
Algorithm	Clustering	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	ΔR_{grp} (%)	$\Delta RMSE(\mu)$ (%)
Original		–	0.000921	0.874940	–	–
ALS	Activity	3	0.000552	0.876573	-40.13%	+0.19%
		5	0.000482	0.875668	-47.68%	+0.08%
		10	0.000358	0.875391	-61.19%	+0.05%
		15	0.000329	0.875398	-64.33%	+0.05%
		20	0.000334	0.875410	-63.74%	+0.05%
		25	0.000317	0.875469	-65.57%	+0.06%

Table 3. R_{grp} and $RMSE$ Metrics: Activity-Based Clustering with **NCF** on **MovieLens 1M**

Activity-Based Strategy for User Grouping $\{G_1, G_2\}$						
Algorithm	Clustering	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	ΔR_{grp} (%)	$\Delta RMSE(\mu)$ (%)
Original		–	0.001717	0.853849	–	–
NCF	Activity	3	0.001359	0.851952	-20.85%	-0.22%
		5	0.001325	0.851555	-22.83%	-0.27%
		10	0.001214	0.851106	-29.31%	-0.32%
		15	0.001080	0.851230	-37.13%	-0.31%
		20	0.001118	0.851143	-34.90%	-0.32%
		25	0.001162	0.851124	-32.33%	-0.32%

Table 4. R_{grp} and $RMSE$ Metrics: Gender-Based Clustering with **ALS** on **MovieLens 1M**

Gender-Based Strategy for User Grouping $\{G_1, G_2\}$						
Algorithm	Clustering	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	ΔR_{grp} (%)	$\Delta RMSE(\mu)$ (%)
Original		–	0.003145	0.884963	–	–
ALS	Gender	3	0.002492	0.892091	-20.78%	+0.81%
		5	0.002232	0.893634	-29.04%	+0.98%
		10	0.002004	0.895262	-36.29%	+1.16%
		15	0.001890	0.896107	-39.91%	+1.26%
		20	0.001885	0.896220	-40.07%	+1.27%
		25	0.001863	0.896476	-40.75%	+1.30%

Table 5. R_{grp} and $RMSE$ Metrics: Gender-Based Clustering with **NCF** on **MovieLens 1M**

Gender-Based Strategy for User Grouping $\{G_1, G_2\}$						
Algorithm	Clustering	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	ΔR_{grp} (%)	$\Delta RMSE(\mu)$ (%)
Original		–	0.002712	0.867263	–	–
NCF	Gender	3	0.002110	0.868539	-22.19%	+0.15%
		5	0.001904	0.869791	-29.82%	+0.29%
		10	0.001764	0.870867	-34.95%	+0.42%
		15	0.001717	0.871155	-36.70%	+0.45%
		20	0.001664	0.871766	-38.63%	+0.52%
		25	0.001653	0.871931	-39.07%	+0.54%

Table 6. R_{grp} and $RMSE$ Metrics: Age-Based Clustering with **ALS** on **MovieLens 1M**

Age-Based Strategy for User Grouping $\{G_1, G_2, G_3, G_4, G_5, G_6, G_7\}$						
Algorithm	Clustering	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	ΔR_{grp} (%)	$\Delta RMSE(\mu)$ (%)
Original		–	0.002170	0.881594	–	–
ALS	Age	3	0.001957	0.888475	-9.82%	+0.78%
		5	0.001909	0.889672	-12.03%	+0.92%
		10	0.001750	0.890907	-19.37%	+1.06%
		15	0.001736	0.891773	-20.01%	+1.15%
		20	0.001700	0.891100	-21.68%	+1.08%
		25	0.001673	0.891400	-22.91%	+1.11%

Table 7. R_{grp} and $RMSE$ Metrics: Age-Based Clustering with **NCF** on **MovieLens 1M**

Age-Based Strategy for User Grouping $\{G_1, G_2, G_3, G_4, G_5, G_6, G_7\}$						
Algorithm	Clustering	h	$R_{grp}(\mu)$	$R_{grp}(\sigma)$	ΔR_{grp} (%)	$\Delta RMSE(\mu)$ (%)
Original		–	0.001367	0.862939	–	–
NCF	Age	3	0.001185	0.863384	-13.34%	+0.05%
		5	0.001113	0.863816	-18.63%	+0.10%
		10	0.001055	0.864861	-22.84%	+0.22%
		15	0.001013	0.865083	-25.96%	+0.25%
		20	0.001003	0.867910	-26.65%	+0.46%
		25	0.001040	0.869375	-23.95%	+0.63%

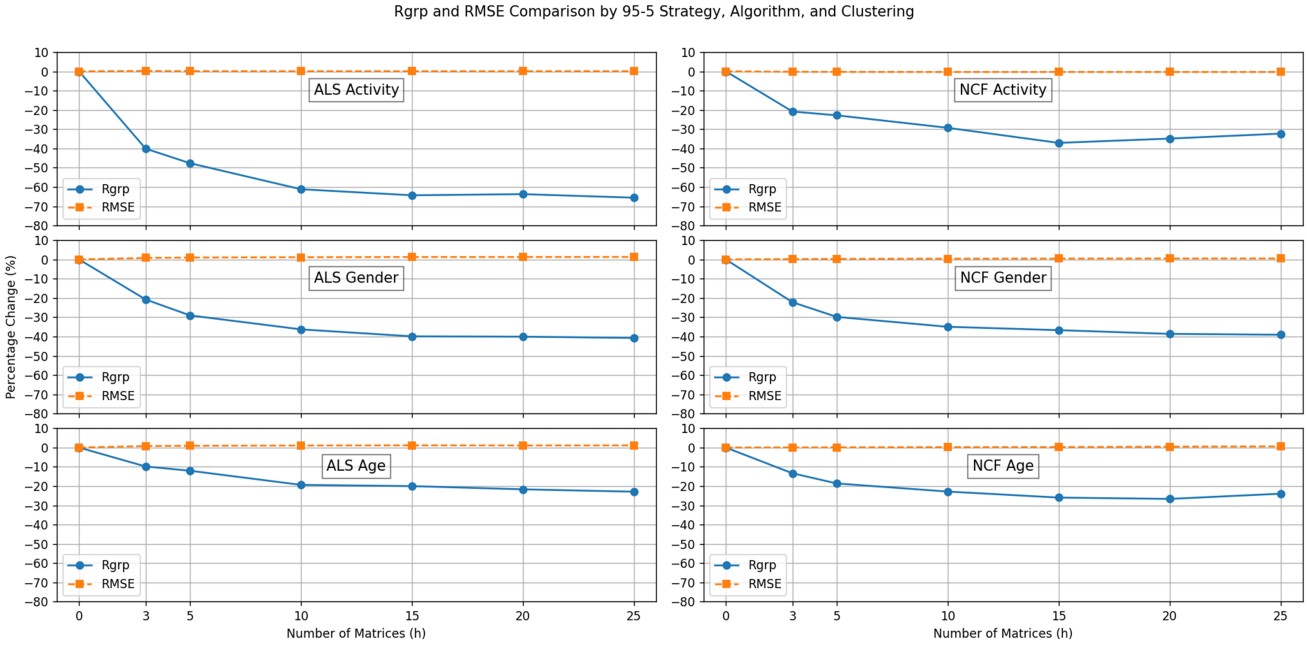

Figure 2. Percentage Reductions of Group Unfairness R_{grp} by Clustering Strategy and Algorithm

illustrating that ALS consistently outperforms NCF in mitigating R_{grp} in all tested clustering strategies. However, regarding the increase in $RMSE$, NCF maintains recommendation accuracy better than ALS.

Furthermore, the figure elucidates an important observation regarding the influence of the parameter h. It becomes evident that the trajectory of the R_{grp} and $RMSE$ metrics tends to stabilize beyond a certain h threshold, indicating a potential equilibrium point where further increases in h do not substantially affect the fairness-accuracy trade-off. Generally, this equilibrium plateau in group unfairness reduction can be observed when h is between 10 and 15, emphasizing the importance of optimal h selection to balance fairness and accuracy in recommendations.

The comparative analysis also featuring the varying susceptibility of clustering strategies to algorithmic interventions aimed at reducing R_{grp}. Specifically, the distinct performance patterns in the clustering strategies by activity level, gender, and age reveal the intricate relationship between data characteristics and the algorithm's efficiency in achieving recommendation fairness. This insight underscores the need for specific strategies for each type of clustering in the implementation of fairness-enhancing algorithms.

Moreover, the minimal impact on $RMSE$ across both algorithms and all clustering strategies reinforces the feasibility of integrating fairness with minimal compromise on recommendation quality. This observation challenges the prevailing notion of an inevitable trade-off between recommendation fairness and accuracy, suggesting that judicious algorithm selection and parameter tuning can indeed align fairness and efficiency goals.

Essentially, Figure 2 not only corroborates the superior capability of the ALS algorithm in diminishing group unfairness but also accentuates the critical role of tailored parameter optimization and algorithmic flexibility in navigating the complexities of equitable recommendation systems. The analysis invites further exploration into adaptive strategies that holistically address the dual objectives of fairness and accuracy in diverse recommendation scenarios.

Through Figure 3, we have a detailed visual representation that contrasts the impact of ALS (*Alternating Least Squares*) and NCF (*Neural Collaborative Filtering*) algorithms on fairness and accuracy in recommendation systems. Using heatmaps for this analysis, the figure effectively illustrates how each algorithm, on average, affects the reduction of group unfairness (R_{grp}) and the increase in Root Mean Squared Error (RMSE) across three different grouping strategies: activity level, gender, and age. This graphical representation facilitates an immediate understanding of the varied effects that different algorithms can have on fundamental aspects of fair

and accurate recommendations. For the calculation of displayed values, we consider the average of reductions obtained for each configuration of the estimated matrices number (h), specifically for values 3, 5, 10, 15, 20, and 25.

The displayed heatmaps allow for a clear and direct reading of the calculated averages for the focused metrics, organized according to algorithms and grouping strategies. The color palette in the heatmaps not only visually differentiates the values but also highlights the differences in results between various combinations of algorithms and grouping strategies, allowing for the intuitive identification of emerging patterns. On the left, we see the representation of the average reduction in group unfairness, and on the right, the graph focuses on the average increase in RMSE. Together, these graphs provide a comprehensive view of the balance between promoting fairness and maintaining precision in the recommendations produced.

The chosen graphic design aims to facilitate a multifaceted comparison between complex variables, offering a visually accessible and immediately comprehensible way to evaluate the performance of different algorithms in varied contexts. This visual approach is essential for guiding informed decision-making in the selection and fine-tuning of recommendation algorithms, seeking to achieve the best possible balance between fairness and accuracy in the proposed systems.

Figure 3 presents the consolidation of average results from experiments with different values of h, providing a complementary and consistent view with observations made earlier in Figure 2. The use of the average for analysis does not distort the overall trends observed in the data, remaining true to the behavior of ALS and NCF algorithms in terms of fairness and precision in recommendations. The color tones in the heatmap quadrants facilitate the identification of differences and similarities between the algorithms and datasets, reinforcing the conclusions obtained in individual analyses.

In the subplot related to the average reduction of group unfairness (R_{grp}), it is observed that results vary significantly based on the groupings and algorithms. The ALS algorithm, when applied to activity level and gender groupings, shows a consistent reduction in unfairness, indicated by more uniform colors in the heatmap quadrants. This suggests that ALS may effectively leverage the characteristics of these groupings to mitigate unfairness in recommendations. On the other hand, when applied to the age grouping, ALS shows a less uniform reduction, possibly due to heterogeneous data distribution in this aspect. Meanwhile, the NCF algorithm, while showing promising results in activity level and gender groupings, exhibits greater variation in unfairness reduction when applied to the age grouping, indicating additional challenges in adapting the model to this specific data dimension.

Regarding RMSE values, a balance is observed in the results for different algorithms across the datasets studied. This indicates that, despite variations in unfairness reduction, the precision of recommendations remains relatively stable. This stability contrasts with the behavior observed in other datasets, where there is a more pronounced trade-off between fairness and precision, accentuating the importance of careful algorithm selection and parameter adjustment to achieve an ideal balance between these two objectives.

In summary, Figure 3, by consolidating the averages of experiments with ALS and NCF in the new activity level, gender, and age groupings, not only confirms the trends observed individually in Figure 2, but also enriches the analysis by emphasizing subtle and significant differences between the groupings and algorithms used.

Figure 4 comprises two radar charts that provide a comparative visualization of the performance of the ALS and NCF algorithms across three distinct user groupings: Activity, Gender, and Age. The first chart focuses on the maximum reduction in group unfairness (R_{grp}), while the second addresses the maximum increase in Root Mean Squared Error (RMSE). These charts allow for an immediate visual analysis of each algorithm's capability to promote fairness, measured by the reduction in R_{grp}, and their effectiveness, observed through the variation in RMSE, under different grouping contexts. The categories on the charts correspond to the analyzed user groupings, providing a basis for direct comparison between the algorithms and demonstrating their relative performances in terms of fairness and accuracy. This representation aims to facilitate understanding of the trade-offs involved in selecting algorithms for recommendation systems that strive to balance precision and fairness.

The radar charts in Figure 4 illustrate the superior capability of ALS in reducing group unfairness across the Activity, Gender, and Age groupings, maintaining an important balance with modest increases in RMSE. ALS demonstrates consistent reductions in group unfairness, with a relatively controlled increase in RMSE across different groupings. This balanced performance denotes the algorithm's efficacy in navigating the trade-off between fairness and precision.

Conversely, the NCF algorithm exhibits more diverse responses to fairness interventions across different groupings. While it contributes to reducing group unfairness, the extent of its effectiveness varies, suggesting less uniform adaptability to the characteristics of the groupings. This observation emphasizes the importance of context-sensitive strategies to optimize the balance between fairness and precision.

The convergence of performance metrics around $h = 10$ indicates the ALS algorithm's ability to achieve an optimal balance between fairness and precision. This convergence suggests a critical threshold beyond which further adjustments in h have diminishing impacts on both fairness and precision. Such insights suggest the importance of parameter optimization to achieve equitable recommendations without compromising precision.

Essentially, Figure 4 reinforces the nuanced dynamics between fairness and precision in recommendation systems, accentuating the importance of algorithmic adaptability and parameter tuning to achieve optimal outcomes across diverse user groupings.

6 Conclusion

In this study, we explored the impact of different algorithms on the development and evaluation of a fairness algorithm in recommendation systems, focusing on grouping users by activity level, gender, and age. We observed that the fairness algorithm effectively reduced Group Unfairness (R_{grp}) in all configurations tested, with significant decreases achieved in

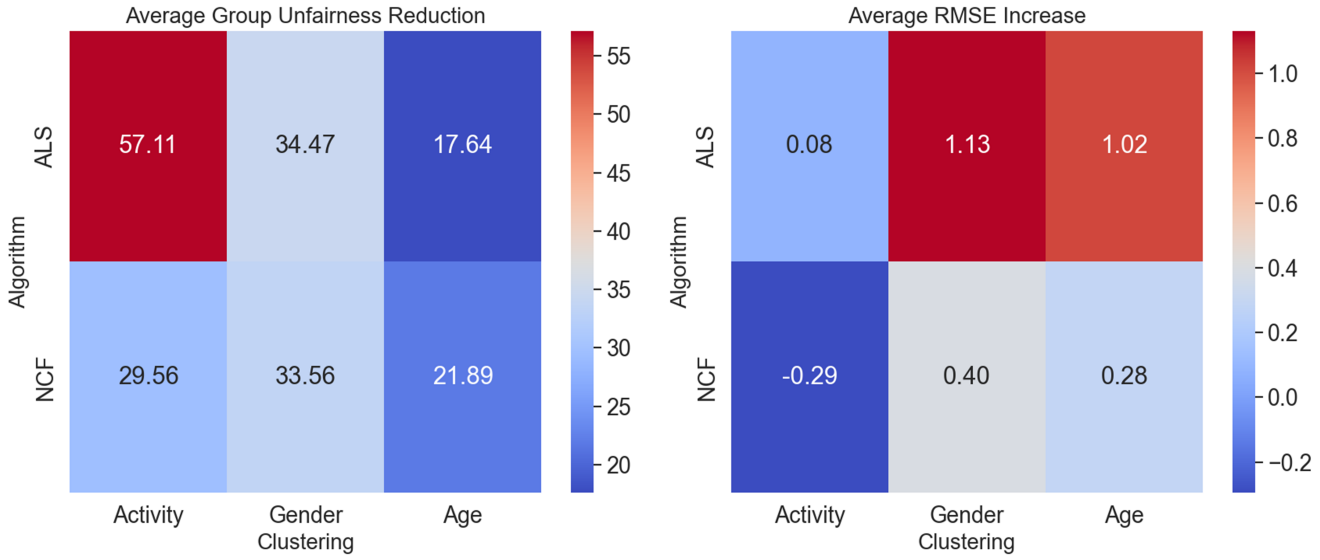

Figure 3. Percentage Reductions of Group Unfairness R_{grp}

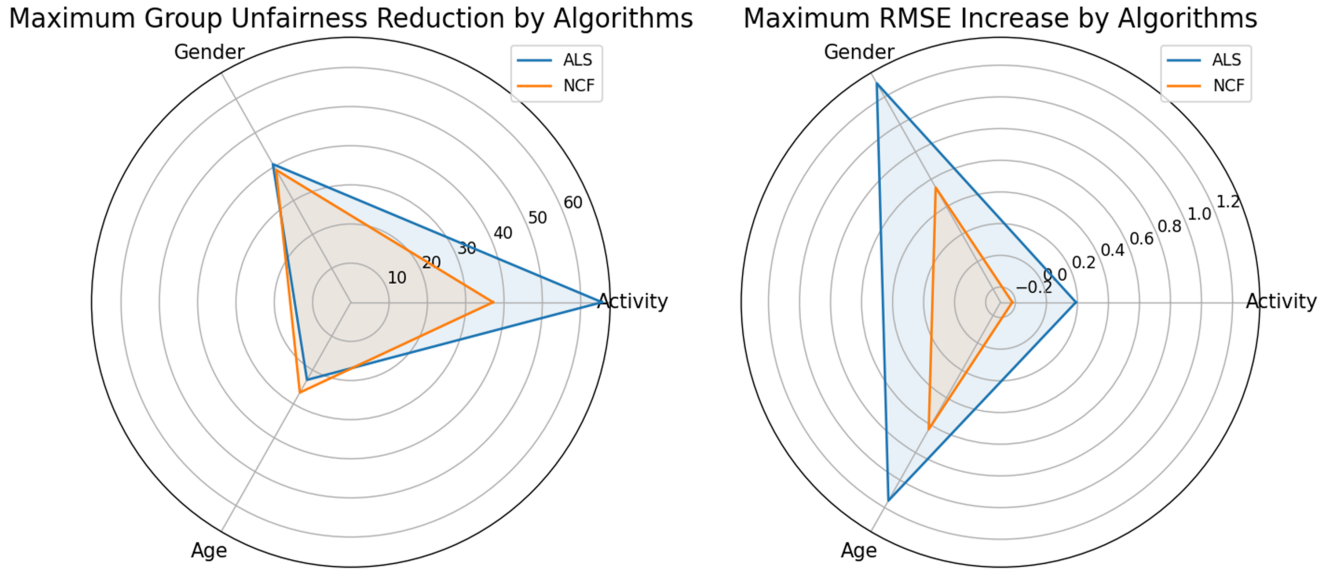

Figure 4. Reductions in Group Unfairness and Increases in RMSE

the MovieLens 1M dataset. Specifically, for $h = 25$ in the grouping strategy by activity level, using the ALS algorithm, a reduction of 65.57% in R_{grp} was noted.

It is important to note that the Root Mean Squared Error (RMSE) did not show a significant increase as a result of these reductions, indicating that the effectiveness of the recommendations remained stable, even with efforts to promote greater fairness. This observation is essential as it demonstrates the possibility of achieving a fairer recommendation system without compromising the quality of the recommendations provided.

Furthermore, a convergence of the fairness algorithm to an optimal number of estimated matrices (h) situated between 10 and 15 was identified. This convergence suggests that adjustments within this h range may offer the best balance between promoting fairness and maintaining precision in the recommendations.

Therefore, the findings of this study illustrate the ability to effectively incorporate considerations of fairness into recommendation systems through the appropriate selection of algorithms and parameter adjustments, without significant detriments to the accuracy of the recommendations. This represents an important step towards the development of recommendation systems that not only respond to user preferences but also promote fairness and equal opportunities among different user groups.

Moreover, we highlight that the proposed framework can be extended to recommendation systems based on implicit feedback, such as Bayesian Personalized Ranking (BPR), through appropriate adaptations. Specifically, the matrix Z could be reconstructed using individual user performance metrics, such as AUC or MRR, instead of prediction errors. The objective function of minimizing the variance between group errors could be preserved, now considering the variance of the group mean performance metrics. In this way, it would be possible to maintain the essence of the methodology while respecting the specific characteristics of systems based on implicit feedback.

Regarding incremental and interactive recommenders, the framework can be applied through periodic or continuous calculation of fairness metrics, allowing dynamic adjustments to the model as new interactions are registered.

For future research, it is planned to evaluate the fairness algorithm in different contexts, utilizing diverse datasets and exploring new user grouping criteria. Additionally, the algorithm will be applied to recommendation strategies beyond classic approaches, such as ALS and NCF, with the objective of assessing its generalization and effectiveness. Furthermore, there are plans to adapt the algorithm to the context of federated learning, aiming to enhance user data privacy. Finally, the framework will be expanded to include a functionality that allows users to select the solution matrix, presenting various outcome options generated from multiple recommendation algorithms during the initial estimation stage. This enhancement will enable users to explicitly prioritize either the promotion of fairness or the maintenance of recommendation accuracy.

Declarations

Authors' Contributions

Rafael V. M. S. was responsible for writing the original draft, research, methodology application, data collection and curation, as well as conducting the formal analysis and translating this work into English. Giovanni V. C. supervised the project, contributing to data analysis and providing insights into the results, as well as editing, correcting, and suggesting improvements for this document. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We would like to thank Federal Institute of Espírito Santo (IFES) and Federal University of Espírito Santo (Ufes) for their academic and financial support.

Availability of data and materials

The datasets used, as well as all implementations and results, are available in the repository: `https://github.com/ravarmes/recsys-fairness`.

References

Beutel, A., Chi, E. H., Cheng, Z., Pham, H., and Anderson, J. (2017). Beyond globally optimal: Focused learning for improved recommendations. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*. DOI: 10.1145/3038912.3052713.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg. Book.

Burke, R., Sonboli, N., and Ordonez-Gauger, A. (2018). Balanced neighborhoods for multi-sided fairness in recommendation. In *FAT*. Available at:`https://proceedings.mlr.press/v81/burke18a.html`.

Dandekar, P., Goel, A., and Lee, D. (2013). Biased assimilation, homophily and the dynamics of polarization. *Proceedings of the National Academy of Sciences of the United States of America*, 110. DOI: 10.1073/pnas.1217220110.

Deldjoo, Y., Anelli, V. W., Zamani, H., *et al*. (2021). A flexible framework for evaluating user and item fairness in recommender systems. *User Modeling and User-Adapted Interaction*, 31:457–511. DOI: 10.1007/s11257-020-09285-1.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. S. (2011). Fairness through awareness. *CoRR*, abs/1104.3913. DOI: 10.48550/arXiv.1104.3913.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. DOI: 10.1038/nature14539.

Gurobi Optimization, LLC (2024). Gurobi optimizer. Available at:`https://www.gurobi.com`.

Hardt, M. (2013). On the provable convergence of alternating minimization for matrix completion. *CoRR*, abs/1312.0925. DOI: 10.48550/arXiv.1312.0925.

Hardt, M., Price, E., and Srebro, N. (2016). Equality of opportunity in supervised learning. *CoRR*, abs/1610.02413. DOI: 10.48550/arXiv.1610.02413.

Harper, F. M. and Konstan, J. A. (2015). The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19. DOI: 10.1145/2827872.

Hastie, T., Mazumder, R., Lee, J., and Zadeh, R. (2014). Matrix completion and low-rank svd via fast alternating least squares. DOI: 10.48550/ARXIV.1410.2596.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2 edition. DOI: 10.1007/978-0-387-84858-7.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. (2016). Fast matrix factorization for online recommendation with implicit feedback. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 549–558. ACM. DOI: 10.48550/arXiv.1708.05024.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. (2017). Neural collaborative filtering. In *Proceedings of the 26th International Conference on World Wide Web*, pages 173–182. ACM. DOI: 10.1145/3038912.3052569.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer. DOI: 10.25334/q4ht55.

Kamishima, T. and Akaho, S. (2017). Considerations on recommendation independence for a find-good-items task. In *In 11th ACM Conference on Recommender Systems*. DOI: 10.18122/B2871W.

Kamishima, T., Akaho, S., and Asoh, H. (2012). Enhancement of the neutrality in recommendation. In *In Proc. of the 2nd Workshop on Human Decision Making in Recommender Systems*, pages 8–14. Available at:`https://ceur-ws.org/Vol-893/paper2.pdf`.

Kamishima, T., Akaho, S., Asoh, H., and Sakuma, J. (2018). Recommendation independence. In Friedler, S. A. and Wilson, C., editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 187–201. PMLR. Available at:`https://proceedings.mlr.press/v81/kamishima18a.html`.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. In *Communications of the ACM*, volume 60, pages 84–90. ACM. DOI: 10.1145/3065386.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444. DOI: 10.1038/nature14539.

Niemiec, W., Borges, R., and Barone, D. (2022). Artificial intelligence discrimination: how to deal with it? In *Anais do III Workshop sobre as Implicações da Computação na Sociedade*, pages 93–100, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/wics.2022.222604.

Rastegarpanah, B., Gummadi, K. P., and Crovella, M. (2019). Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, WSDM '19. ACM. DOI: 10.1145/3289600.3291002.

Ruback, L., Avila, S., and Cantero, L. (2021). Vieses no aprendizado de máquina e suas implicações sociais: Um estudo de caso no reconhecimento facial. In *Anais do II Workshop sobre as Implicações da Computação na Sociedade*, pages 90–101, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/wics.2021.15967.

Taso, F., Reis, V., and Martinez, F. (2023). Discriminação algorítmica de gênero: Estudo de caso e análise no contexto brasileiro. In *Anais do IV Workshop sobre as Implicações da Computação na Sociedade*, pages 13–25, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/wics.2023.229980.

Wang, H., Wang, N., Yeung, D.-Y., and Yeung, D.-Y. (2018). Collaborative filtering with social regularization. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 135–144. ACM. DOI: 10.1145/3219819.3219823.

Yao, S. and Huang, B. (2017). Beyond parity: Fairness objectives for collaborative filtering. *CoRR*, abs/1705.08804. DOI: 10.48550/arXiv.1705.08804.

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., and Dwork, C. (2013). Learning fair representations. In Dasgupta, S. and McAllester, D., editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 325–333, Atlanta, Georgia, USA. PMLR. Available at:`https://proceedings.mlr.press/v28/zemel13.html`.