Bias in Recommendation Systems: Detection and Mitigation

Mustapha N [1] Ndlovu N [1] Khoza N [1] Ngcobo S [1]

[1]The Council for Scientific and Industrial Research (CSIR)

Introduction

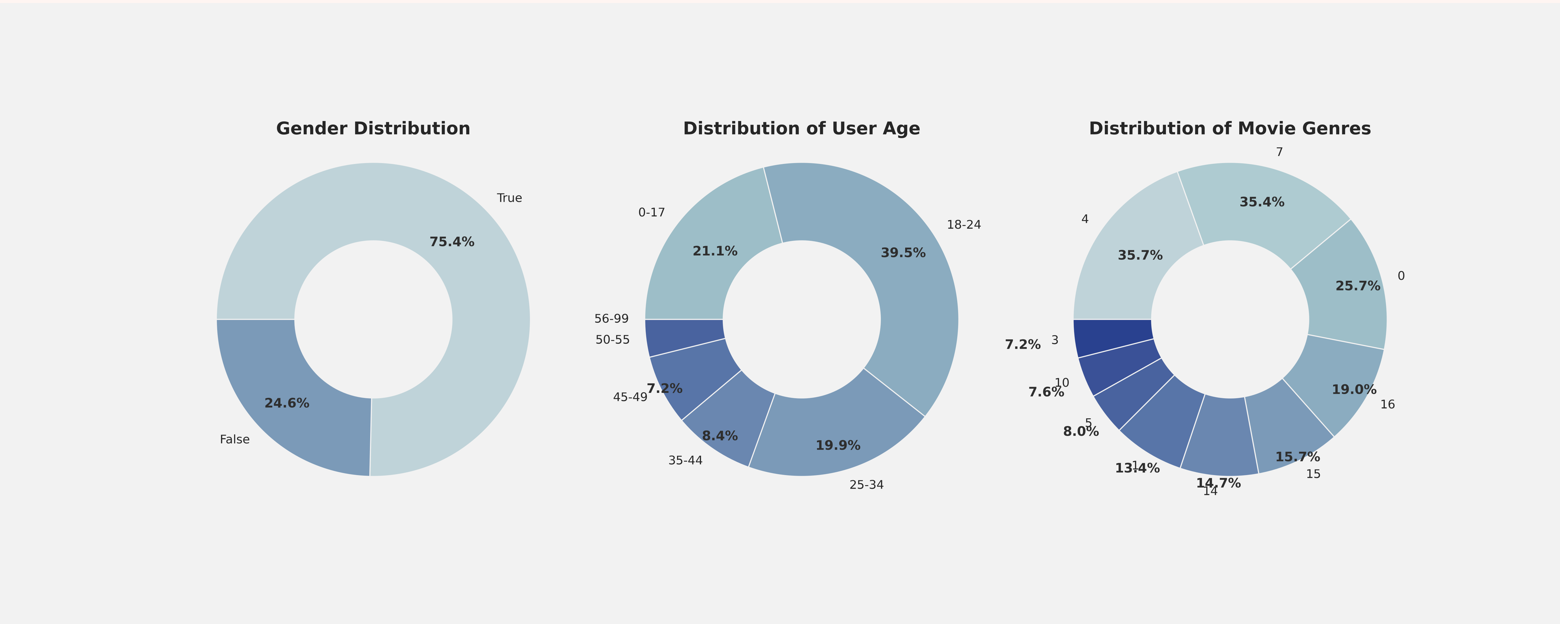
In today's digital era, personalized recommendations have revolutionized the online experience. These systems have become integral to various industries, including e-commerce, and music/video streaming services, enhancing user experience and aligning recommendations with individual interests to achieve this level of customisation relies on user data, which predominantly observational rather than experimental[1].

Biases stem from various sources in the data such as the user demographics, historical preferences, and patterns of interaction within the system. Some of the biases can cause the system to suffer from but not limited to selection bias, position bias, exposure bias, and popularity bias [2].

As these systems expand into diverse domains, such as healthcare and e-learning, and the increasing reliance on personalized recommendations, it becomes crucial to understanding and addressing the biases within these systems and ensure fairness and transparency. This research aims to advocate for fairness and transparency in recommendation systems by identifying and mitigating bias . The poster consist of three main sections:baseline model development,bias detection and mitigation.

Data Imbalances

Imbalances in data can lead to potential bias in recommendation systems due to the unequal distribution of user preferences, item popularity, or other relevant characteristics. The figure below provides insights into the distributions of gender, user age groups, and movie genres within the dataset.These imbalances stand as potential catalysts for bias, shaping the trajectory of recommendations and influencing user experiences



Baseline Model Development

A foundational hybrid model was developed to serve as a benchmark for evaluating subsequent bias techniques.

- Deep Cross Network (DCN) was used to capture low and high-order feature interactions and learn intricate representations.
- Incorporating user features and movie features significantly improved the model expressiveness.

Results

- Training Process: Gradual decrease in Root Mean Square Error (RMSE) and training loss values, indicated effective learning and generalization.
- Training Set Performance: Achieved RMSE of **0.8499** with corresponding loss of **0.7191**, demonstrating good performance on the training set.
- Test Set Performance:RMSE of **0.9542** with corresponding loss of **0.90213**, slightly higher than the training set, but still reasonable.

Conclusion: Despite a minor increase in deviation within the test set, the model demonstrates effective learning and a note-able ability to generalize.

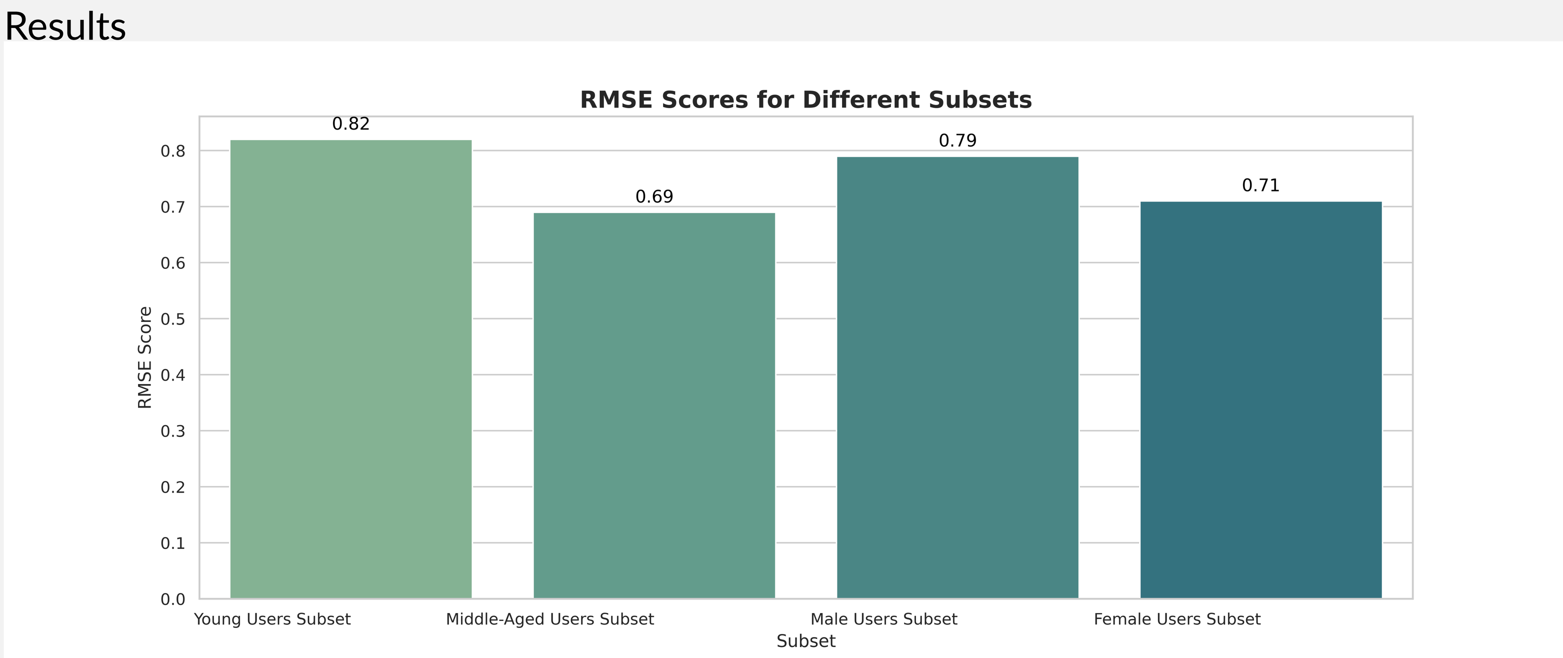
Bias Detection Approach

Two Approaches to Explore Bias were taken:

1. Subset Datasets Based on User Demographics and Ratings:
   - Subsets for specific user population groups (young, middle-aged, male, female) were created.
   - Separate data-sets for positive (ratings ≥ 4) and negative (ratings < 4) examples were created.
   - The models performance was then evaluated on each subset of data.
2. Calculation of Eigenvector Centrality Using Ratings as Weights:
   - The recommendation system was represented as a directed graph (DiGraph).
   - Nodes for users and items were defined,and edges to represent interactions were created, this captured the relationships between users and the items they have interactions with.
   - The eigenvector centrality was calculated for each node in the network.
   - This was done to understand the influence of users or movies in shaping recommendations.

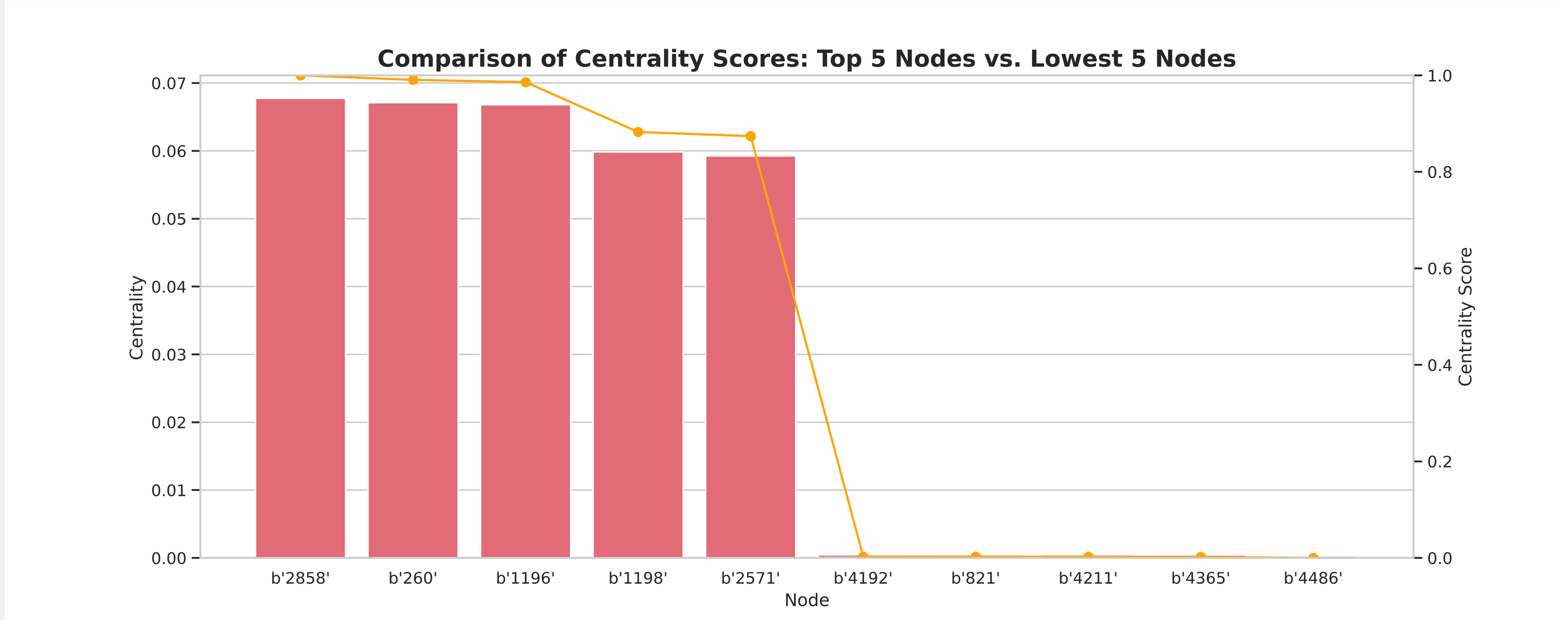
1.Bias Detection:Subset Datasets

Results



The observed variations in RMSE values across different population subsets above indicate performance discrepancies based on user demographics, suggesting potential bias in the recommendation system.

2.Bias Detection:Eigenvector centrality



Eigenvector centrality analysis provides insights into the influential users and popular movies in the recommendation system. Users with higher centrality values have a more significant impact in shaping the recommendations.These findings further underscore the presence of bias in the system.

Bias Mitigation

To mitigate bias in the recommendation system, the popularity weight for each movie based on its average rating was calculated,as it was found that ratings had a significant impact on the recommendation system.

Popularity Weight Let $avg_rating(m)$ be the average rating of a movie m.
The popularity weight $pw(m)$ for the movie m is then calculated as:

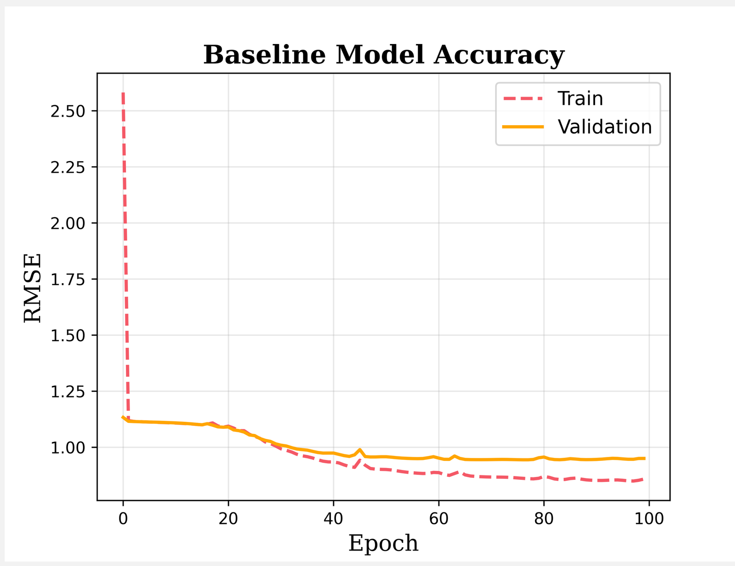
$$pw(m) = \begin{cases} a_vgrating(m) & \text{for popular movies} \\ 1 & \text{for non-popular movies} \end{cases}$$

Regularized Rating Calculation:
The regularized rating for a movie m by a user u is computed as follows:
Let $r(m,u)$ be the rating given by user u to movie m.
Let $pw(m)$ be the popularity weight of movie m.
The regularized rating $rr(m,u)$ is then given by:

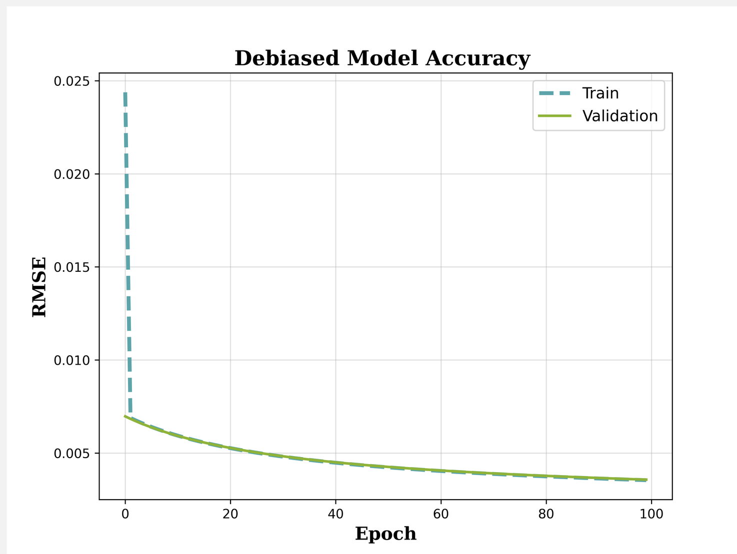
$$rr(m,u) = r(m,u) \times pw(m)$$

A regularization term was also added to the loss function

Effectiveness of Regularization Technique



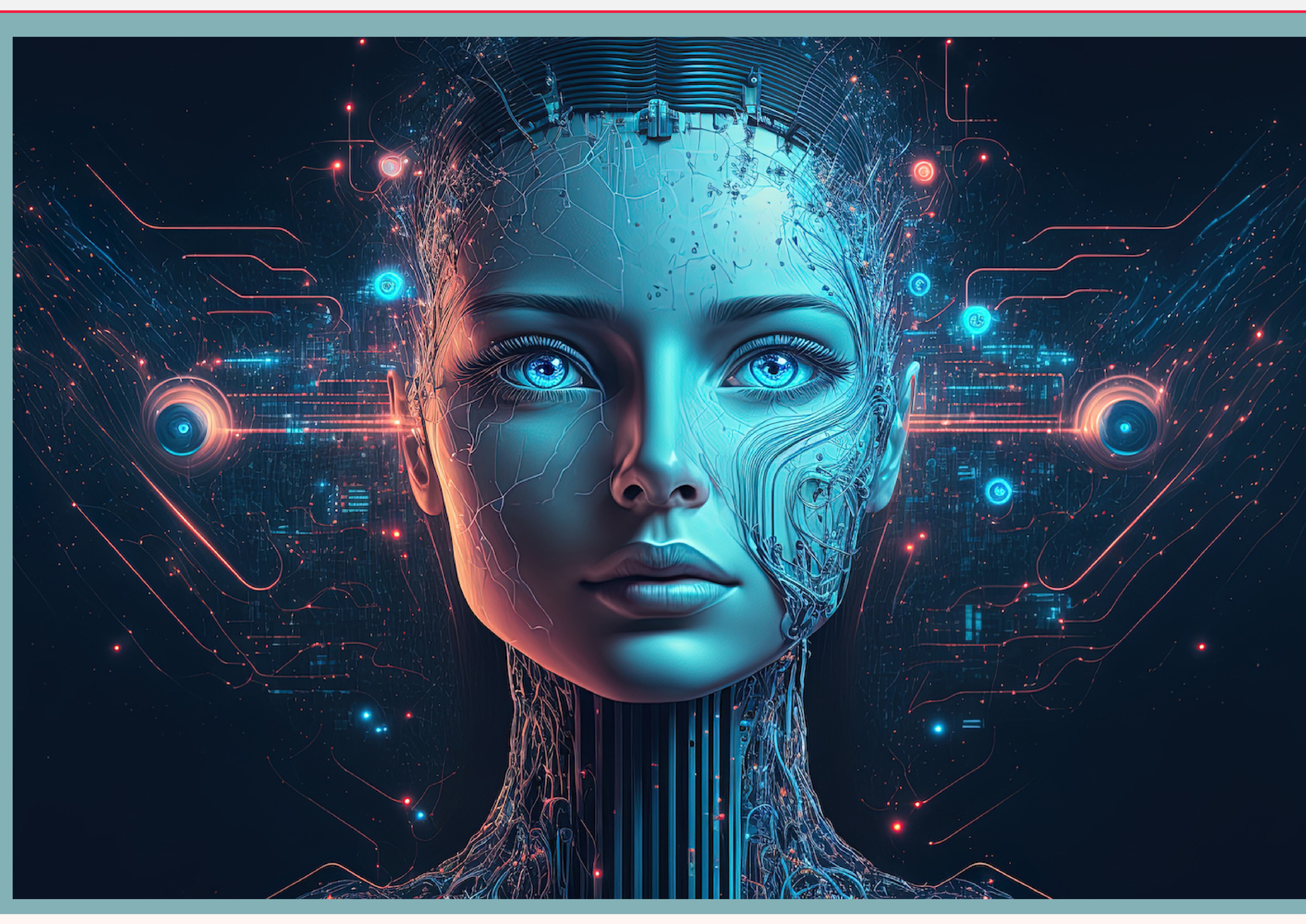
(a) Baseline Model Accuracy



(b) Debiased Model Accuracy

Figure 1. Comparison of Model Accuracy's

The debiased model yielded significantly better results compared to the baseline model



Independent t-test

In addition, an independent t-test compared the performance of the two models. The analysis resulted in a t-statistic of **-13.54** and a p-value of **0.0054**, indicating a significant difference between the two models' performance. The model with popularity weight regularization demonstrated superior performance

Conclusion

Discrepancies in model performance were identified by generating subset data-sets, revealing potential bias .In addition calculating the eigenvector centrality further shed light on the influential users and popular movies have, revealing potential sources of bias within the recommendation system.The developed de-biased model showed exceptional performance with low RMSE values, measuring around **2.2873e-04** and **2.3162e-04** on the test set,ultimately contributing to more equitable and accurate recommendations across diverse user subgroups and creating more fair and transparent recommendation systems.

References

[1] Jiawei Chen et al. "Bias and Debias in Recommender System: A Survey and Future Directions". In: ACM Trans. Inf. Syst. 41.3 (Feb. 2023). issn: 1046- 8188. doi: 10.1145/3564284. url:

Get in touch