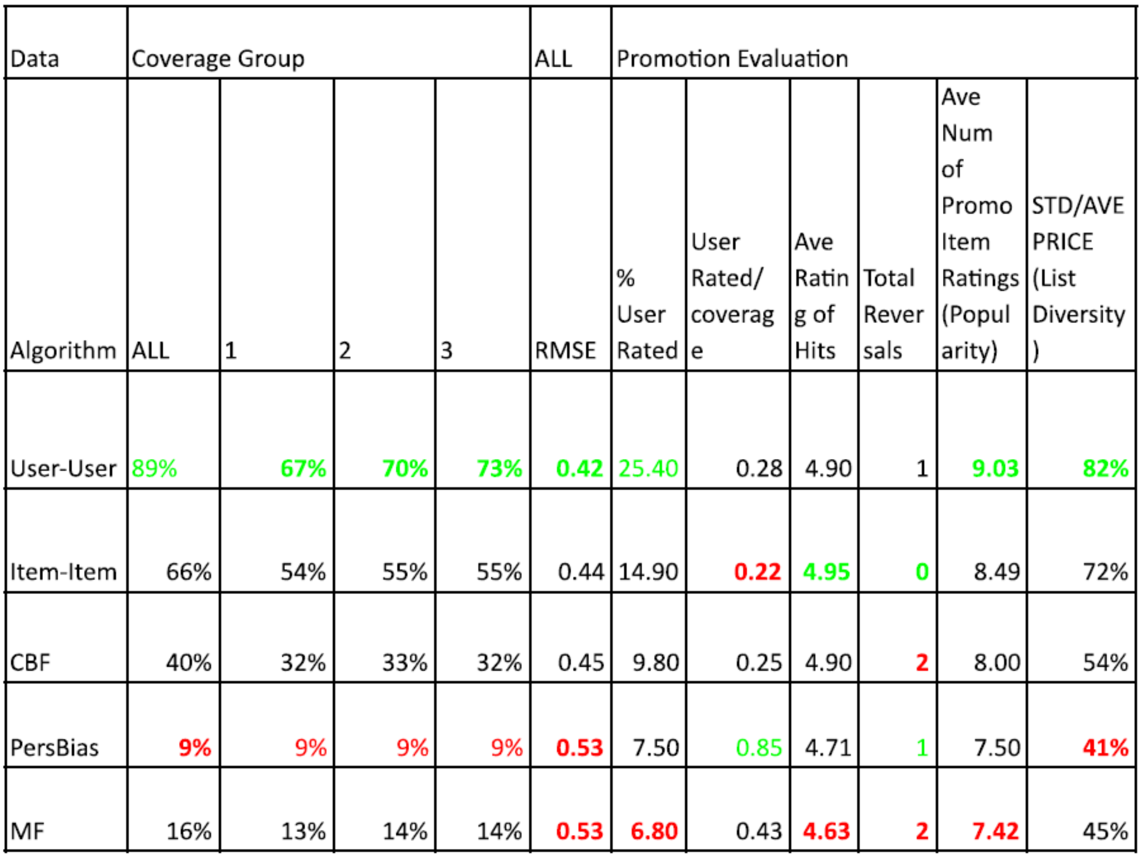
# Mix, Proposal and Reflection

The first two options approach the recommender's performance in terms of how good it predicts the users' ratings, so its only evaluation will be in terms of RMSE.

The third approach have the intuition that, if we get the top 1 recommendation from each algorithm, the resulting 5 item list will have a better performance in terms of identyfing 'good' items to users. In this case, we defined the good items if the recommender suggested an already bought item for an user. Therefore, the final measurement of this hibridization mechanism is through the precision@5, as we end up with a 5 item list.

The final mixing algorithm has the underlying theory of how collaborative filtering mechanisms perform with items that had not enough users/items in its calculations. As a well known weakness of these recommenders, the idea was to check how many items we would affect if we established a threshold of enough data in order for us to use a collaborative filtering. Otherwise, if the item doesn't have enough support in form of users' ratings we could have a support of a content based recommendation, or even, in last case, a non personalised one.

The evaluation results were shown in a table below.



| Data | Coverage Group | | | | ALL | Promotion Evaluation | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Algorithm | ALL | 1 | 2 | 3 | RMSE | % User Rated | User Rated/ coverage | Ave Rating of Hits | Total Reversals | Ave Num of Promo Item Ratings (Popularity) | STD/AVE PRICE (List Diversity) |
| User-User | 89% | 67% | 70% | 73% | 0.42 | 25.40 | 0.28 | 4.90 | 1 | 9.03 | 82% |
| Item-Item | 66% | 54% | 55% | 55% | 0.44 | 14.90 | 0.22 | 4.95 | 0 | 8.49 | 72% |
| CBF | 40% | 32% | 33% | 32% | 0.45 | 9.80 | 0.25 | 4.90 | 2 | 8.00 | 54% |
| PersBias | 9% | 9% | 9% | 9% | 0.53 | 7.50 | 0.85 | 4.71 | 1 | 7.50 | 41% |
| MF | 16% | 13% | 14% | 14% | 0.53 | 6.80 | 0.43 | 4.63 | 2 | 7.42 | 45% |

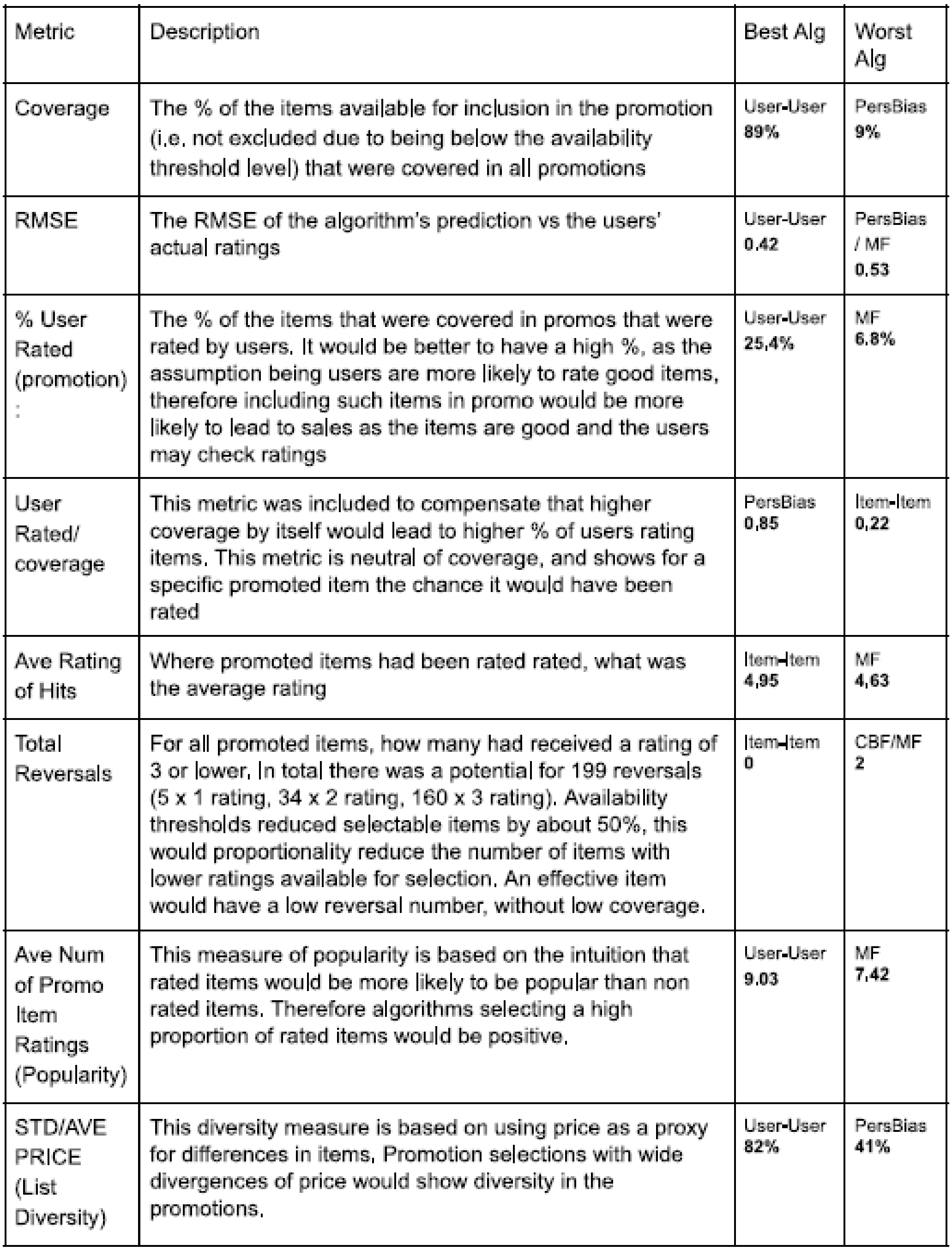
We can see in the cumulative histogram that only around 20% of the rated items had 10 or more ratings. This signals us that maybe we can prioritize the use of a content based recommender or even a non personalised one for the majority of the items which don't have a sufficient amount of ratings in order to make the collaborative filtering algorithms to be stable. Each algorithm has strenghts and weakness. So we can combine them.

Example, blend collaborative filtering with popularity: make sure people get relevant items for them AND try to make sure they will like it

- Means of Hibidrization:
- Combine Items Scores

Linear Blends or feature weighting linear scaling (as more ratings an item has, the less emphasis is put on a content based and more weight to the rank of the collaborative filtering is given for example)

- Combine Item Ranks: Combina based on Output, not score
- Integrated Models
- Advanced Models
  - Conditionally Switch Models
  - Deep Integration (Putting content based computations inside a collaborative filtering)
  - Matrix Factorization on Hibrid Data - SVD++ or SVDFeatures? No ideas..
  - Netflix Challenge - Crazy stuff not real for real life
  - Learning to Rank - FunkSVD for example



| Metric | Description | Best Alg | Worst Alg |
|---|---|---|---|
| Coverage | The % of the items available for inclusion in the promotion (i.e. not excluded due to being below the availability threshold level) that were covered in all promotions | User-User<br>89% | PersBias<br>9% |
| RMSE | The RMSE of the algorithm's prediction vs the users' actual ratings | User-User<br>0.42 | PersBias / MF<br>0,53 |
| % User Rated (promotion) : | The % of the items that were covered in promos that were rated by users. It would be better to have a high %, as the assumption being users are more likely to rate good items, therefore including such items in promo would be more likely to lead to sales as the items are good and the users may check ratings | User-User<br>25,4% | MF<br>6.8% |
| User Rated/ coverage | This metric was included to compensate that higher coverage by itself would lead to higher % of users rating items. This metric is neutral of coverage, and shows for a specific promoted item the chance it would have been rated | PersBias<br>0,85 | Item-Item<br>0,22 |
| Ave Rating of Hits | Where promoted items had been rated rated, what was the average rating | Item-Item<br>4,95 | MF<br>4,63 |
| Total Reversals | For all promoted items, how many had received a rating of 3 or lower. In total there was a potential for 199 reversals (5 x 1 rating, 34 x 2 rating, 160 x 3 rating). Availability thresholds reduced selectable items by about 50%, this would proportionality reduce the number of items with lower ratings available for selection. An effective item would have a low reversal number, without low coverage. | Item-Item<br>0 | CBF/MF<br>2 |
| Ave Num of Promo Item Ratings (Popularity) | This measure of popularity is based on the intuition that rated items would be more likely to be popular than non rated items. Therefore algorithms selecting a high proportion of rated items would be positive. | User-User<br>9.03 | MF<br>7,42 |
| STD/AVE PRICE (List Diversity) | This diversity measure is based on using price as a proxy for differences in items. Promotion selections with wide divergences of price would show diversity in the promotions. | User-User<br>82% | PersBias<br>41% |