Temporal Drift of User Rating On Movie Recommender Systems

Mustafa Katipoğlu

RA. Sercan AYGÜN

Aim Of The Project

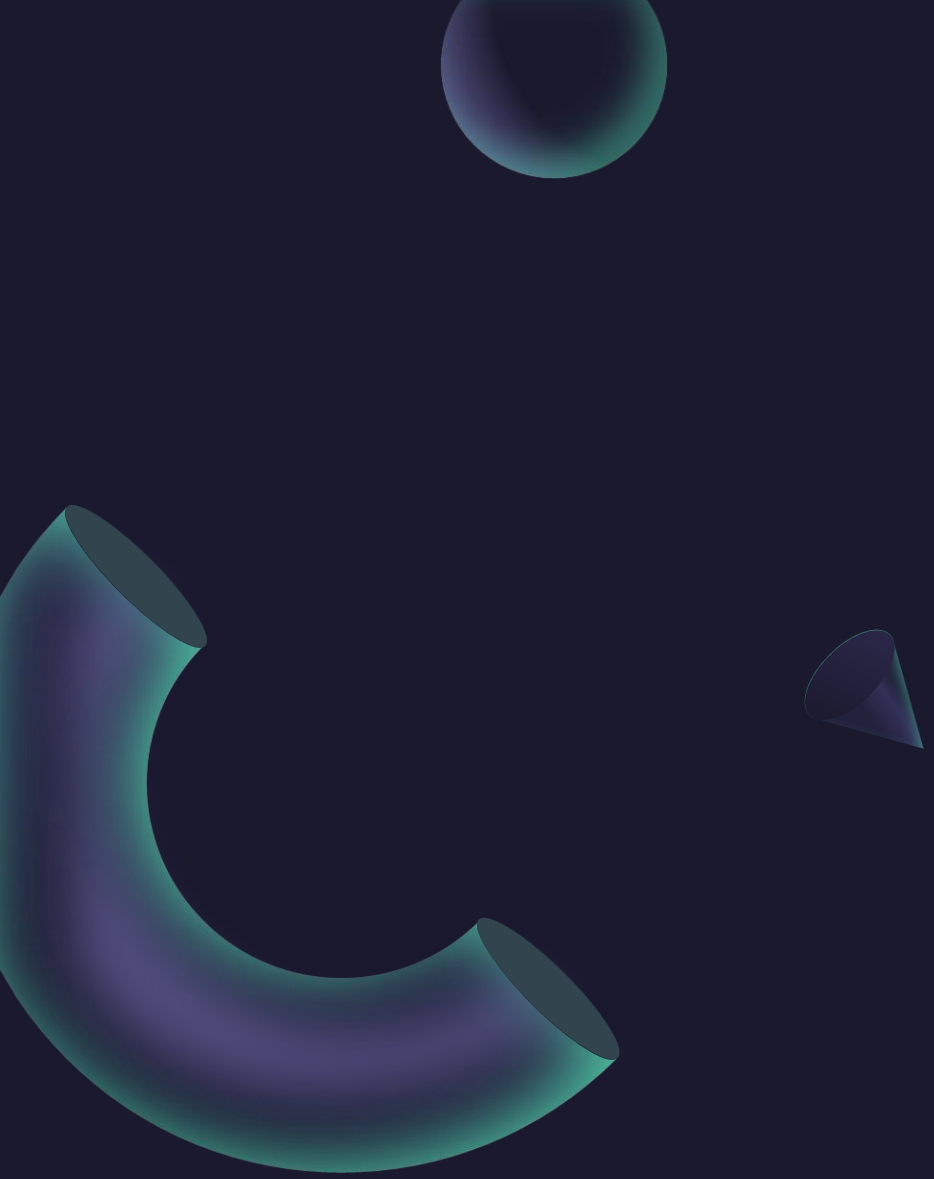
- Creating a new similarity measure that is based on Temporal effects.
- Research on the temporal aspects.

Making Recommendations

- User Based and Item Based Collaborative Filtering(CF).
- User based CF is more personalized.
- Uses similarities between the neighbors.

Pearson Correlation

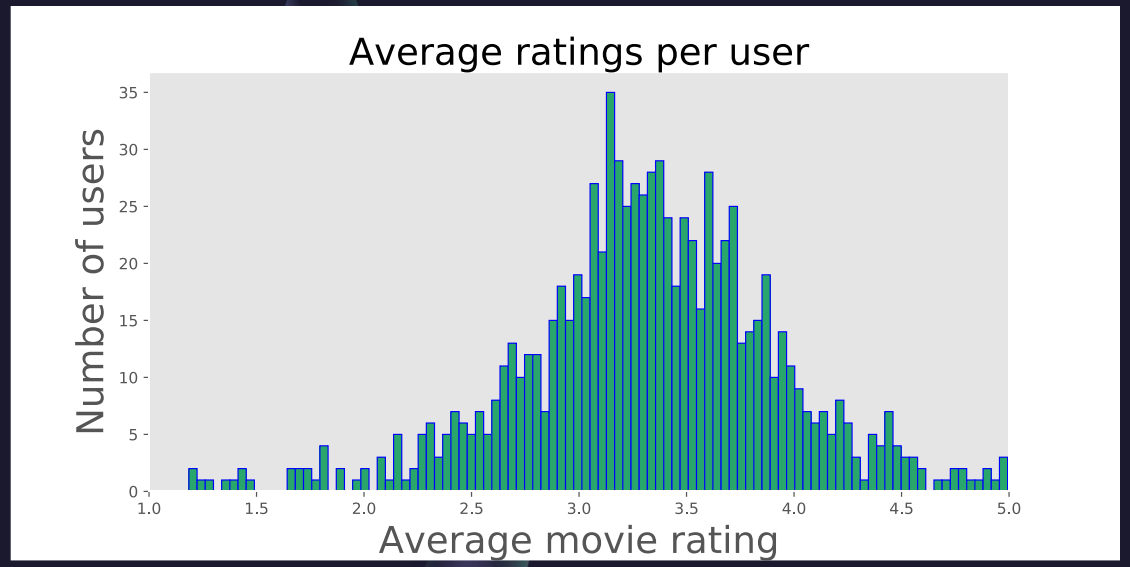
- The classical way of making prediction.
- Based on statistical correlations of the ratings given by the users.

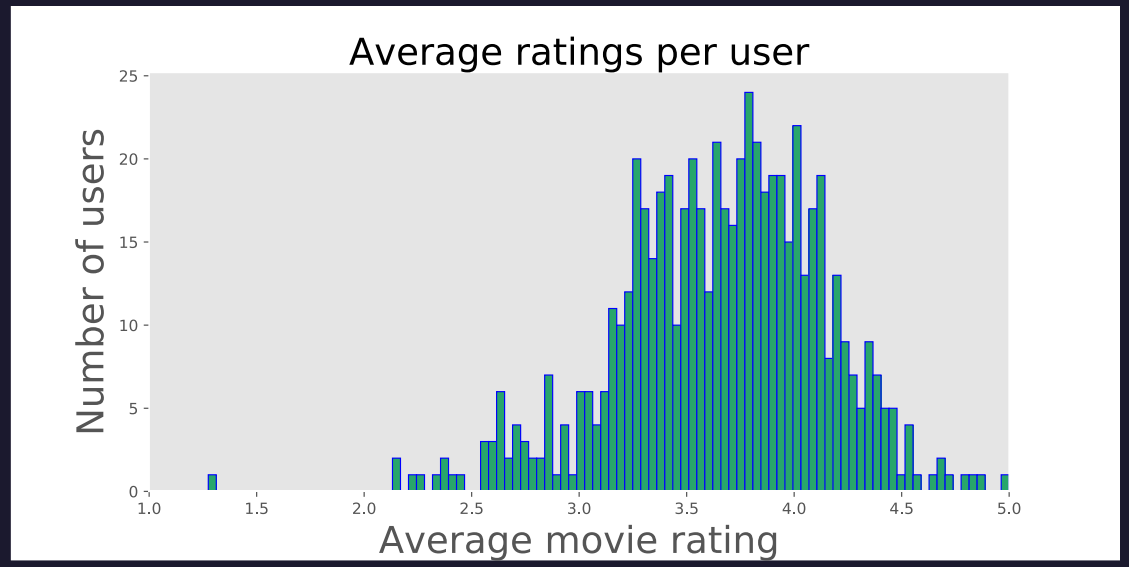


Datasets

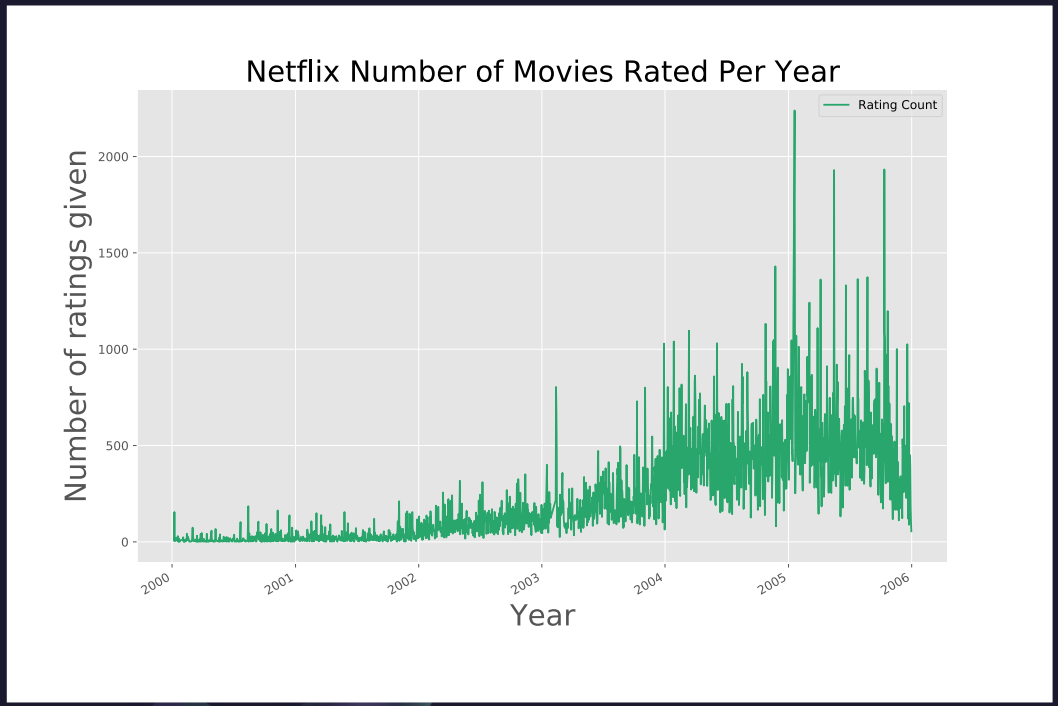
- Movielens 100k Dataset
- Netflix Prize Dataset(Truncated)

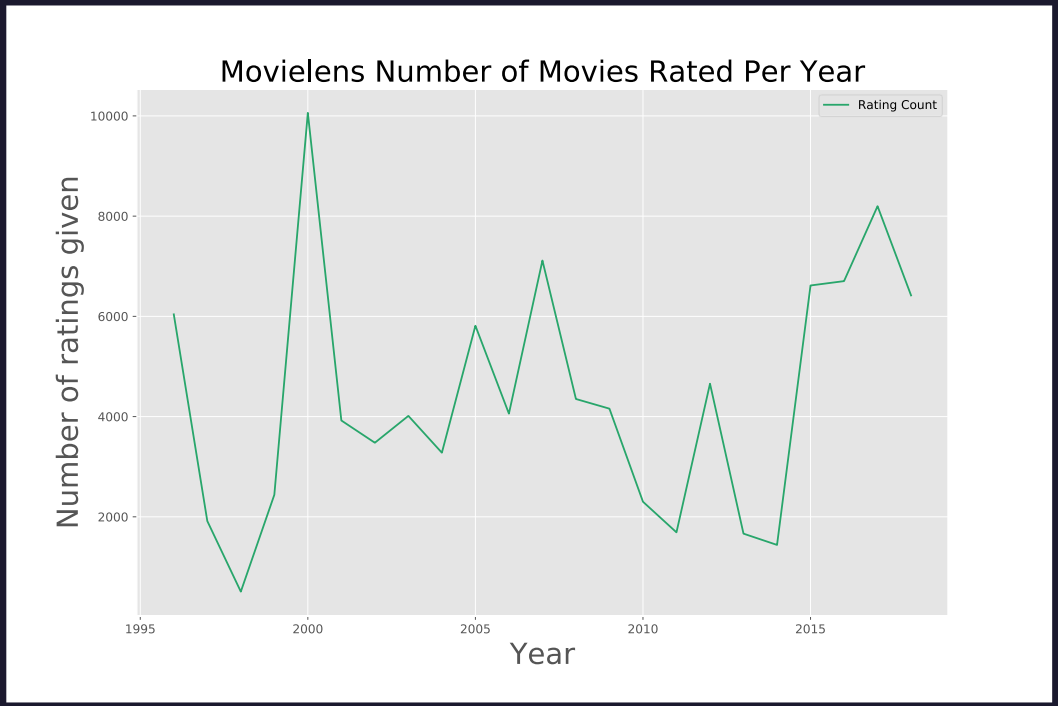
Average User Ratings





Ratings Given Per Year





The Problem

- Temporal aspects is overlooked.
- Drift of user preferences.

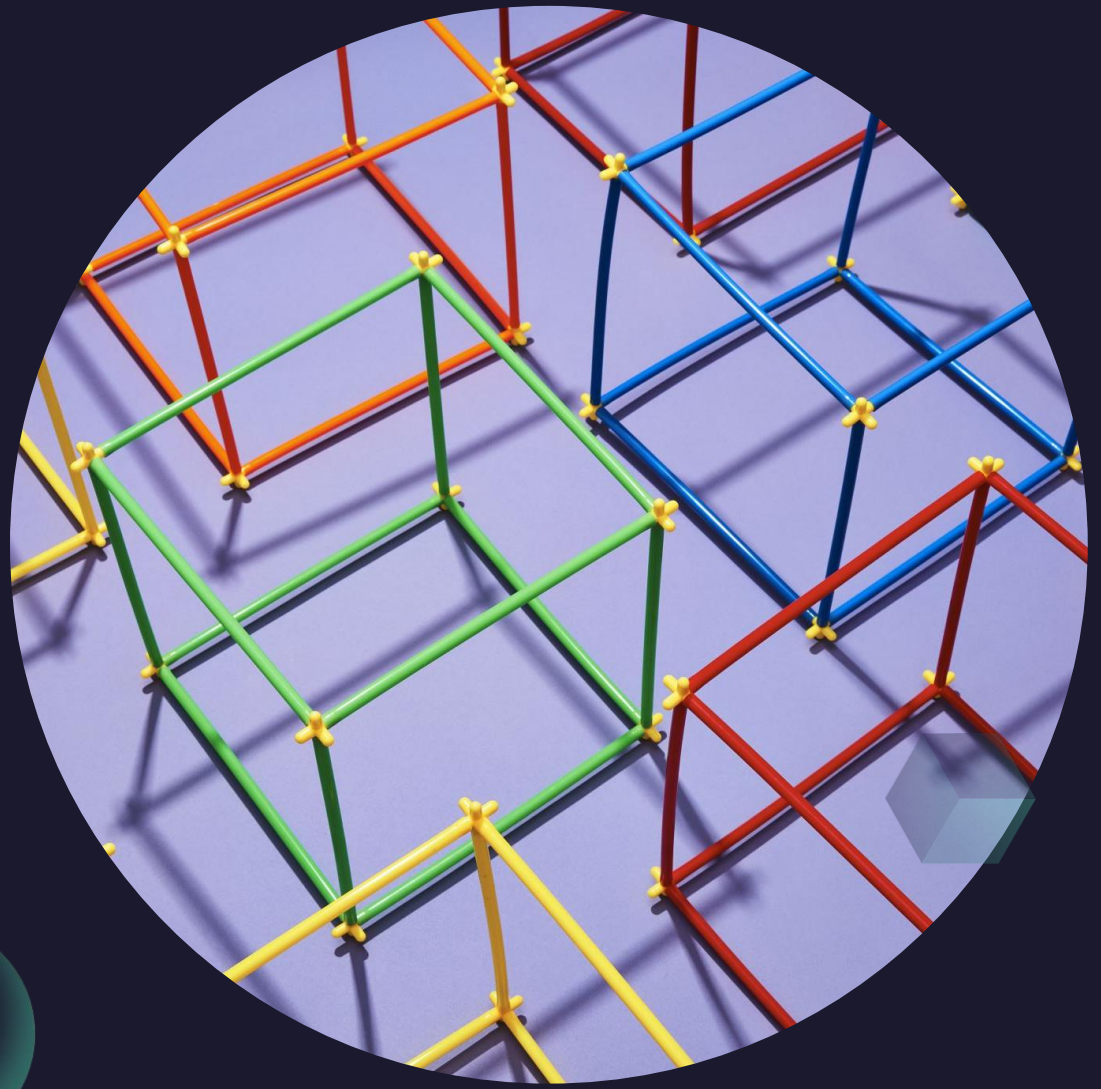
Proposed Approach

Time-bin Based Nearest Neighborhood

- Focus on temporal drift
- Focus on temporal locality

Presented Framework

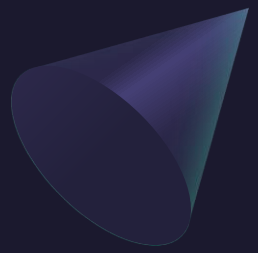
- Written in Python
- Skeleton Framework
- Demand on future work.



Timebin Based Nearest Neighborhood

- Take Time-bins
- Find Similar Time-bins
- Predict as weighted average.

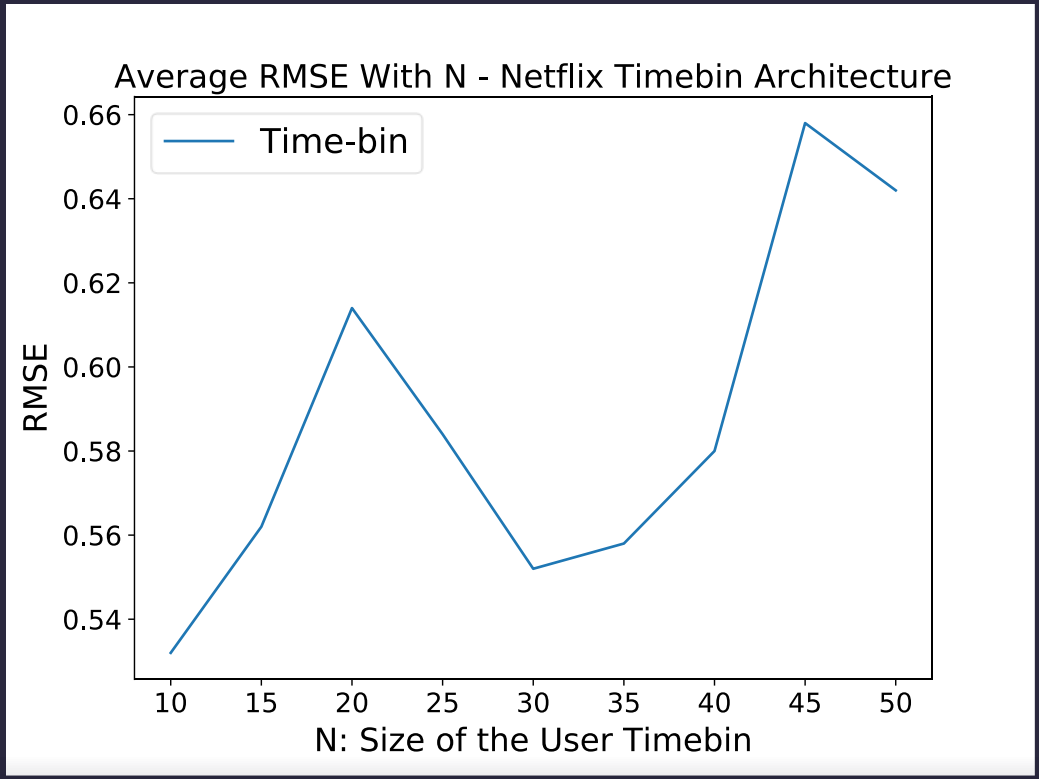
Timebins

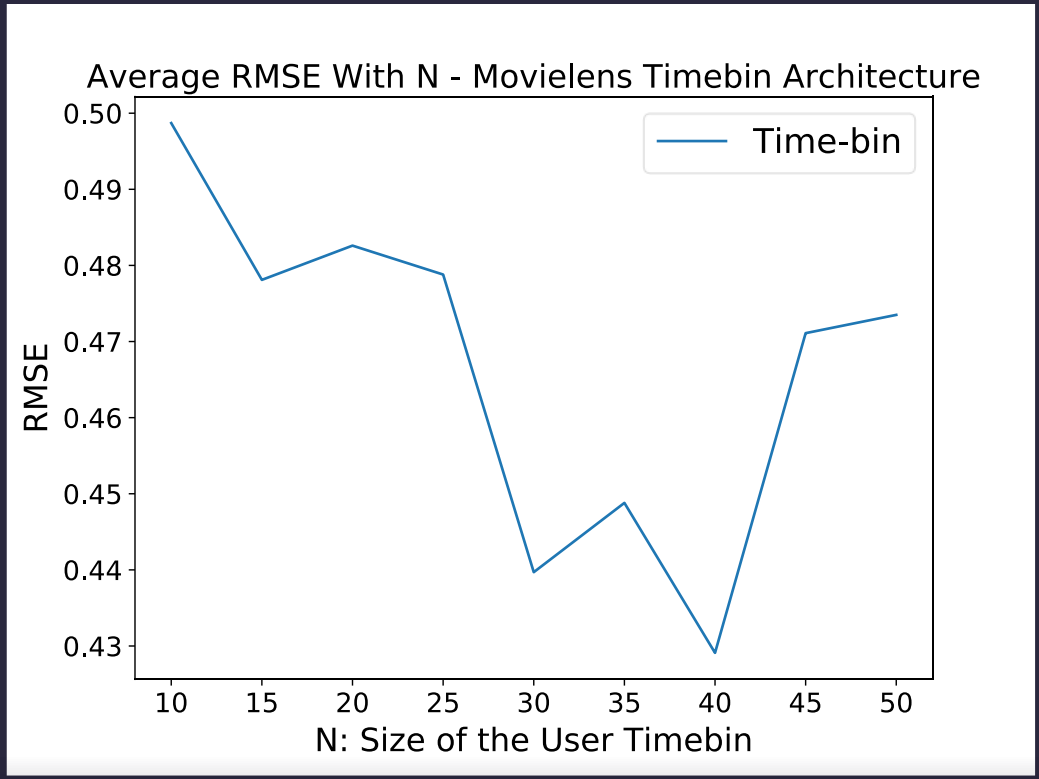


First Way: Static Length of Time – Dynamic Size of Bins

Actual Way: Dynamic Length of Time – Static Size of Bins

Time-bin Size (N)





Constraints

- For Target: At least Time-bin size number of rating history.
- For Neighbors:
  - Need to rate the target movie
  - Need to have at least 2 more common rating

Finding Similar Timebins

- Make sure all the constraints are met
- Use pearson correlation between time-bins only.
- Use the common ratings found in the time-bins.
- Take the avg of the user as the avg of the whole watch history.

Multi Class Classifiacation Based Evaluation

Why ?

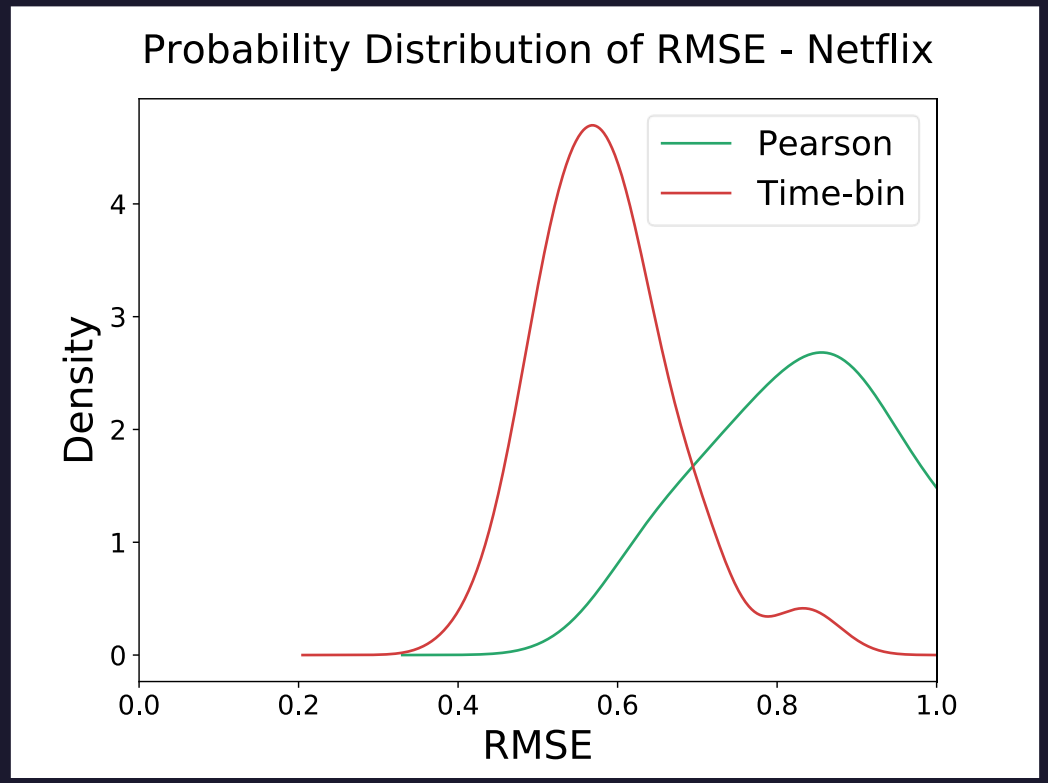
→ Inconsistent datasets.

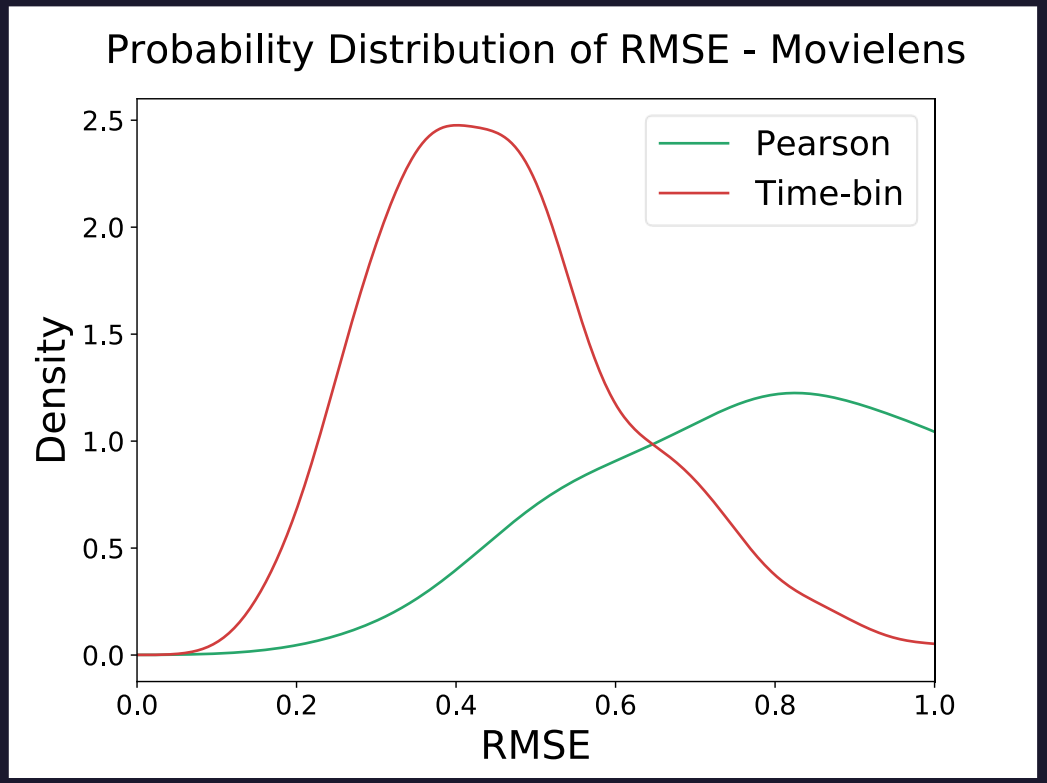
- Netflix 1to 5 - 5 rating class
- Movielens 0.5 to 5 -10 rating class

How ?

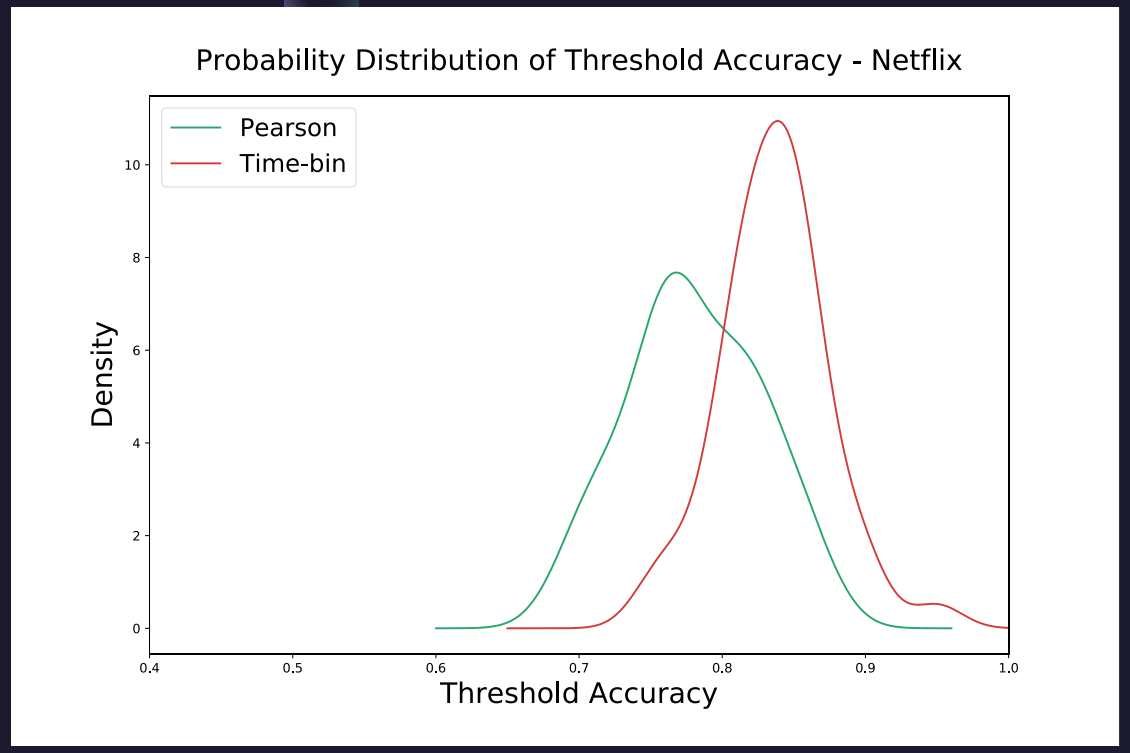
- Suppose there are three classes: C1, C2, and C3
- "TP of C1" is all C1 instances that are classified as C1.
- "TN of C1" is all non-C1 instances that are not classified as C1.
- "FP of C1" is all non-C1 instances that are classified as C1.
- "FN of C1" is all C1 instances that are not classified as C1.
- To find these four terms of C2 or C3 you can replace C1 with C2 or C3.

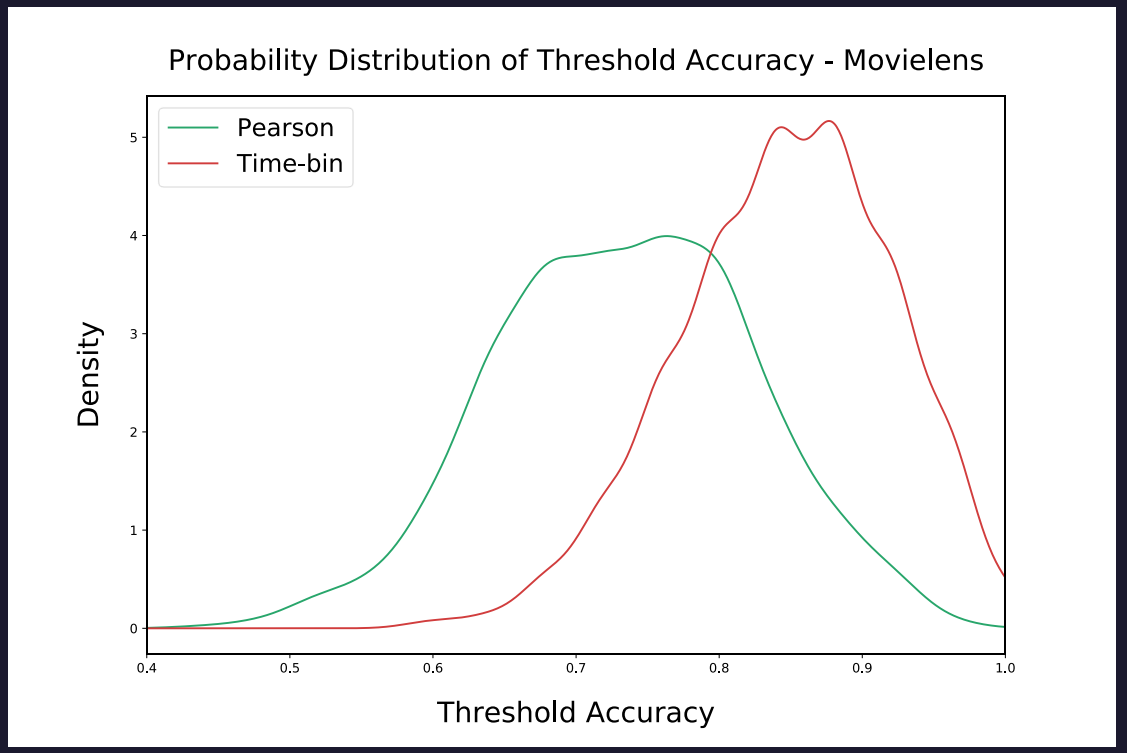
RMSE


Probability Distribution of RMSE - Netflix
Pearson
Time-bin
Density
RMSE
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0

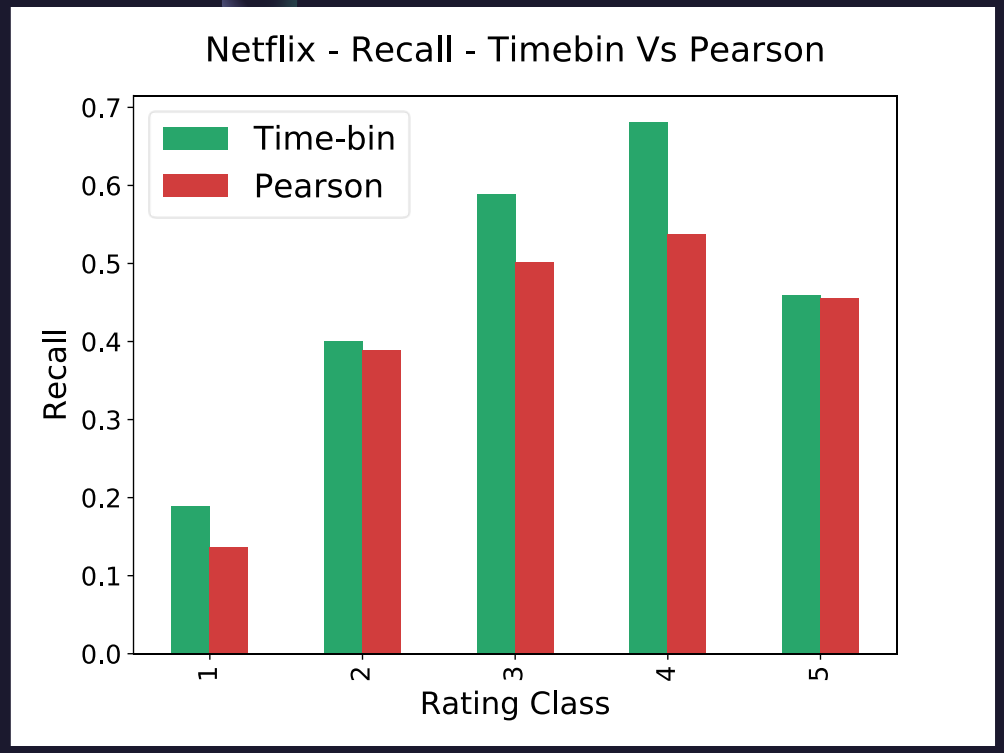

Probability Distribution of RMSE - Movielens
Pearson
Time-bin
Density
RMSE
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.2
0.4
0.6
0.8
1.0

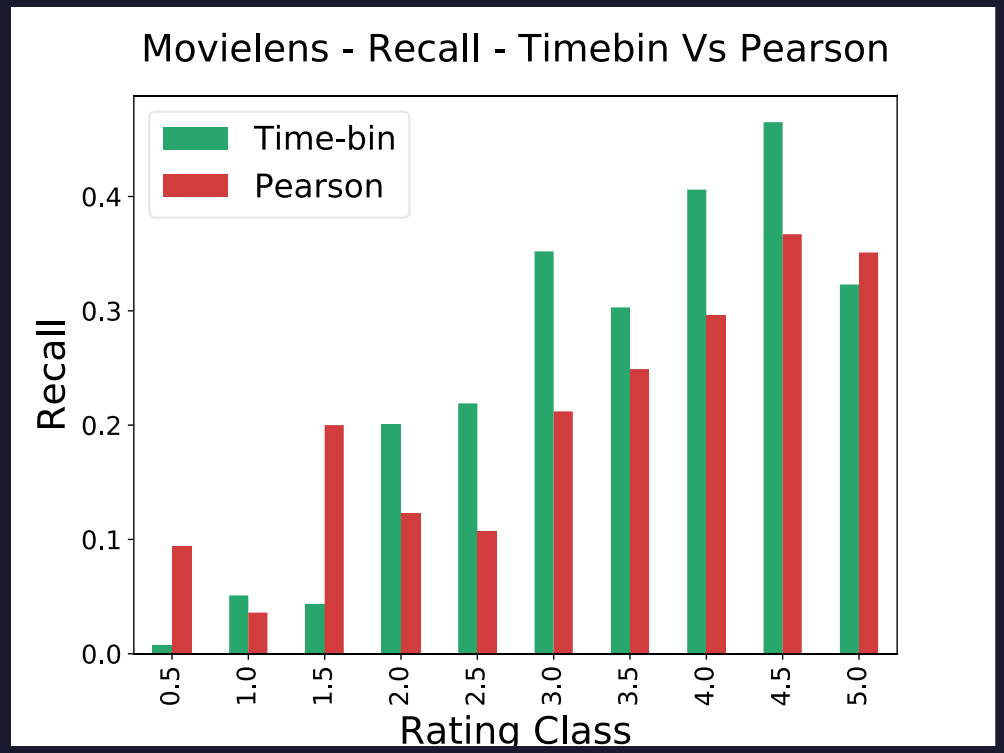
Threshold Accuracy



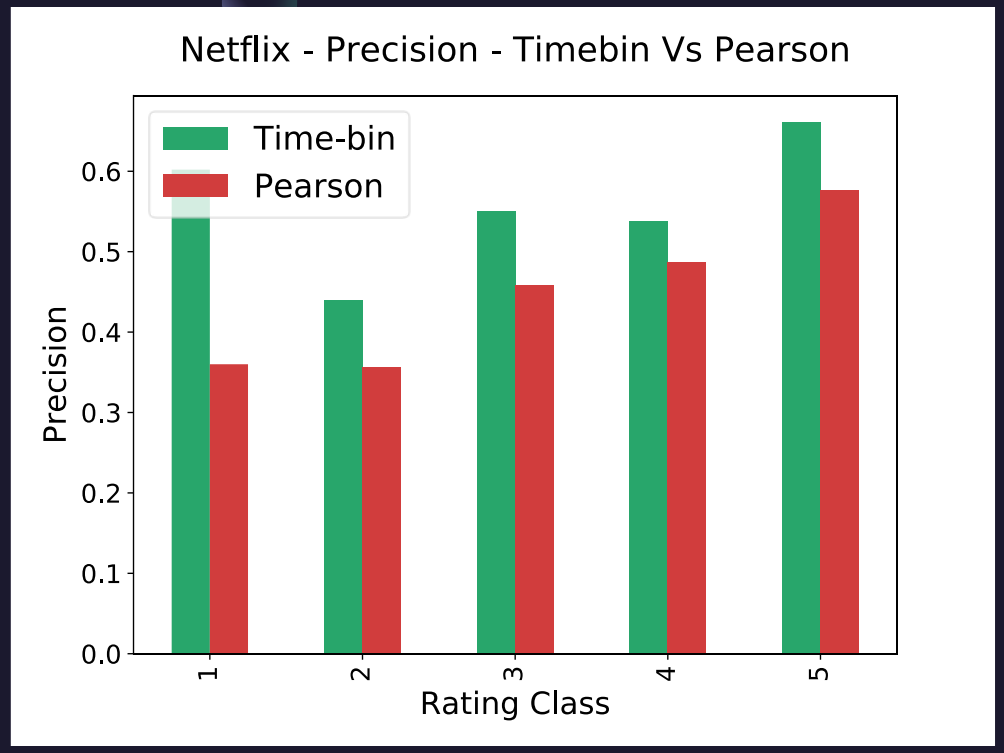


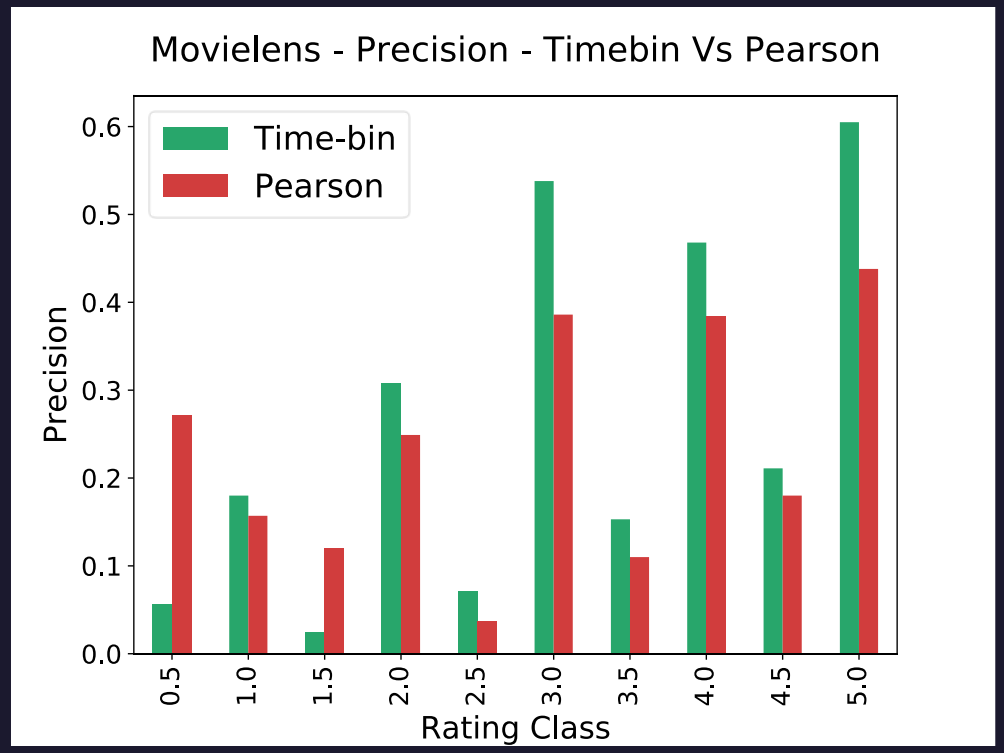
Recall


Netflix - Recall - Timebin Vs Pearson
Time-bin
Pearson
Recall
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1
2
3
4
5
Rating Class

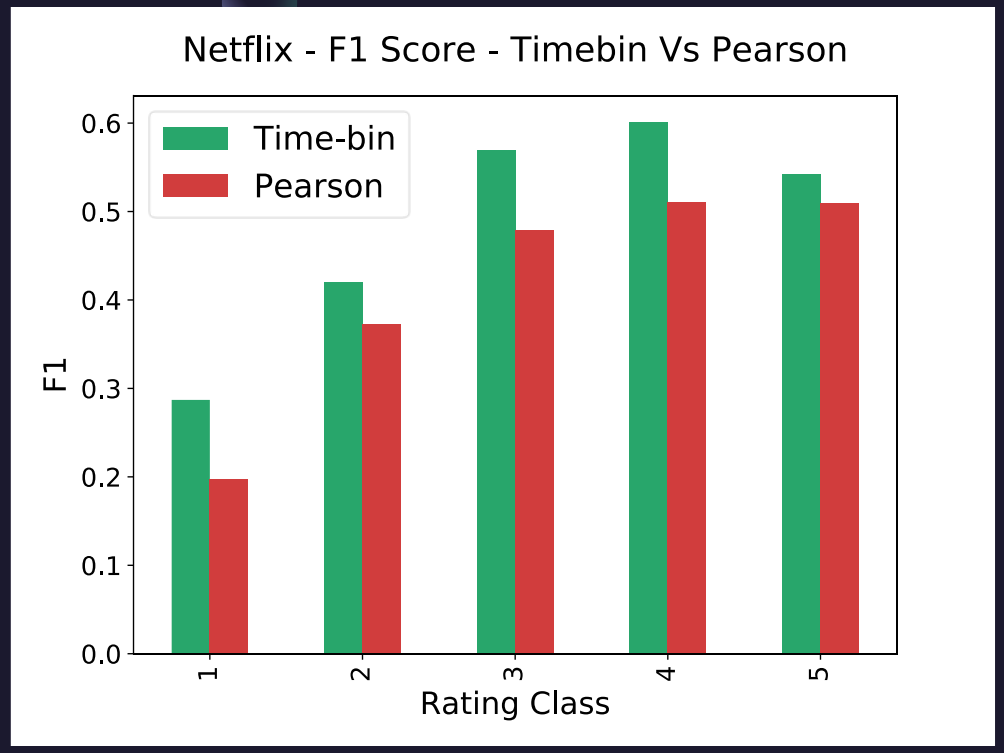

Movielens - Recall - Timebin Vs Pearson
Time-bin
Pearson
Recall
0.0
0.1
0.2
0.3
0.4
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Rating Class

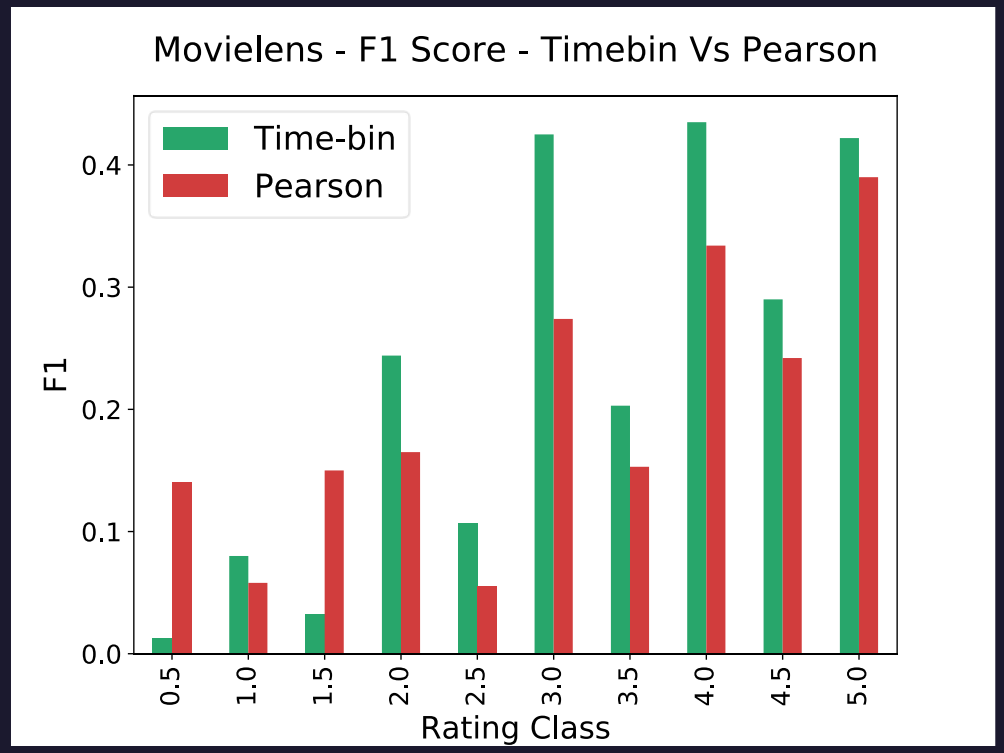
Precision


Netflix - Precision - Timebin Vs Pearson
Time-bin
Pearson
Precision
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1
2
3
4
5
Rating Class

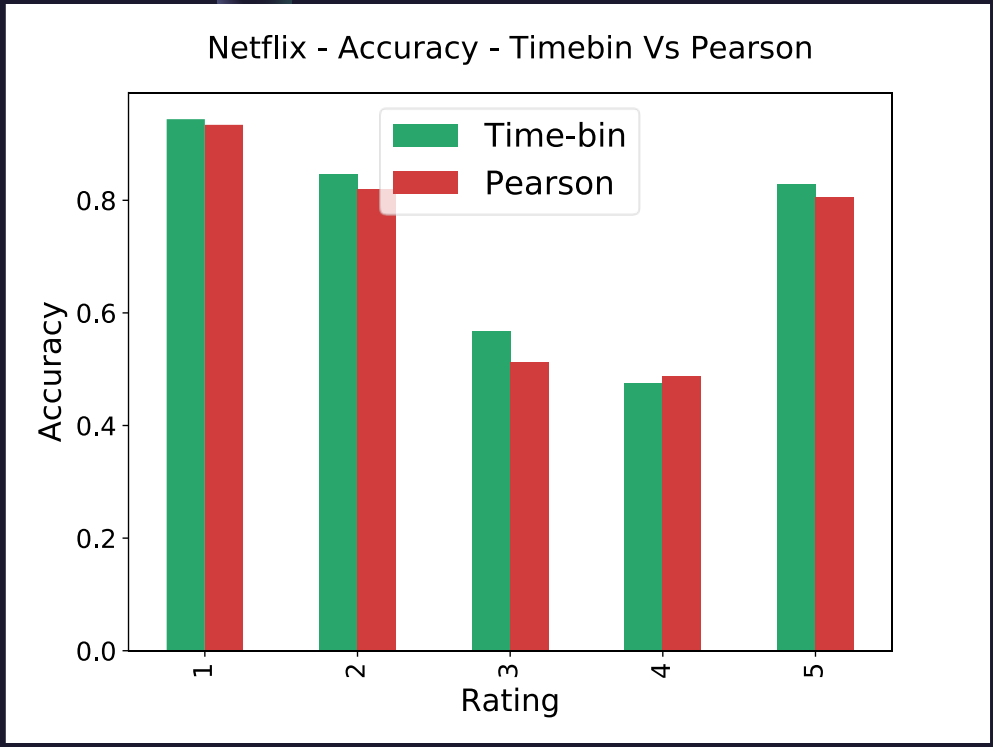

Movielens - Precision - Timebin Vs Pearson
Time-bin
Pearson
Precision
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Rating Class

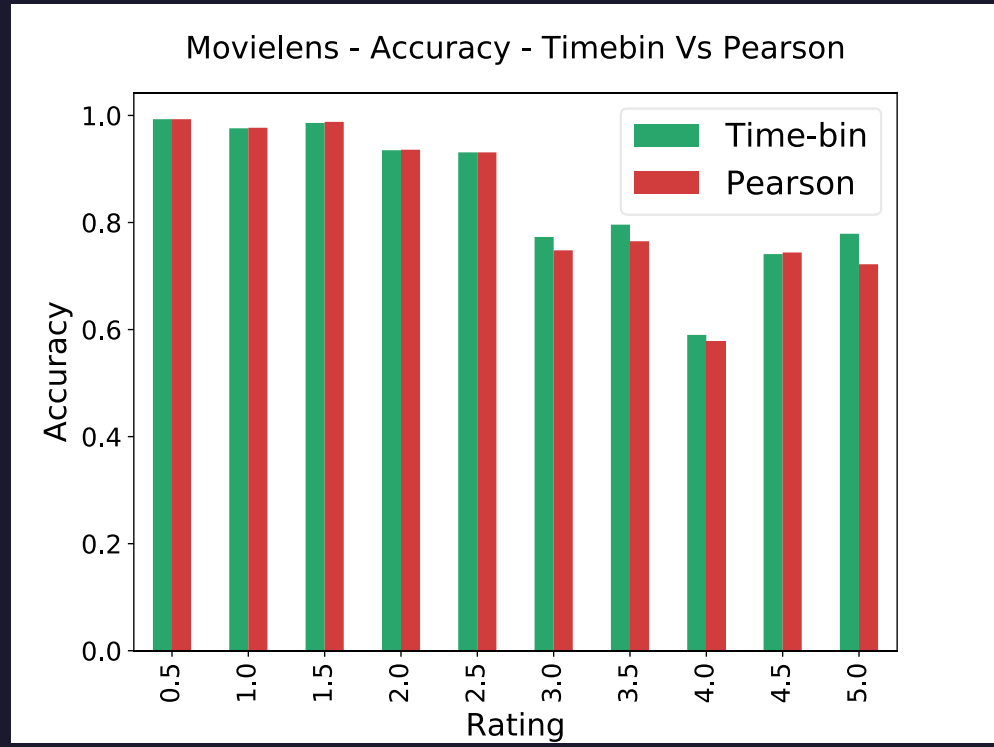
F1 Score


Netflix - F1 Score - Timebin Vs Pearson
Time-bin
Pearson
F1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
1
2
3
4
5
Rating Class

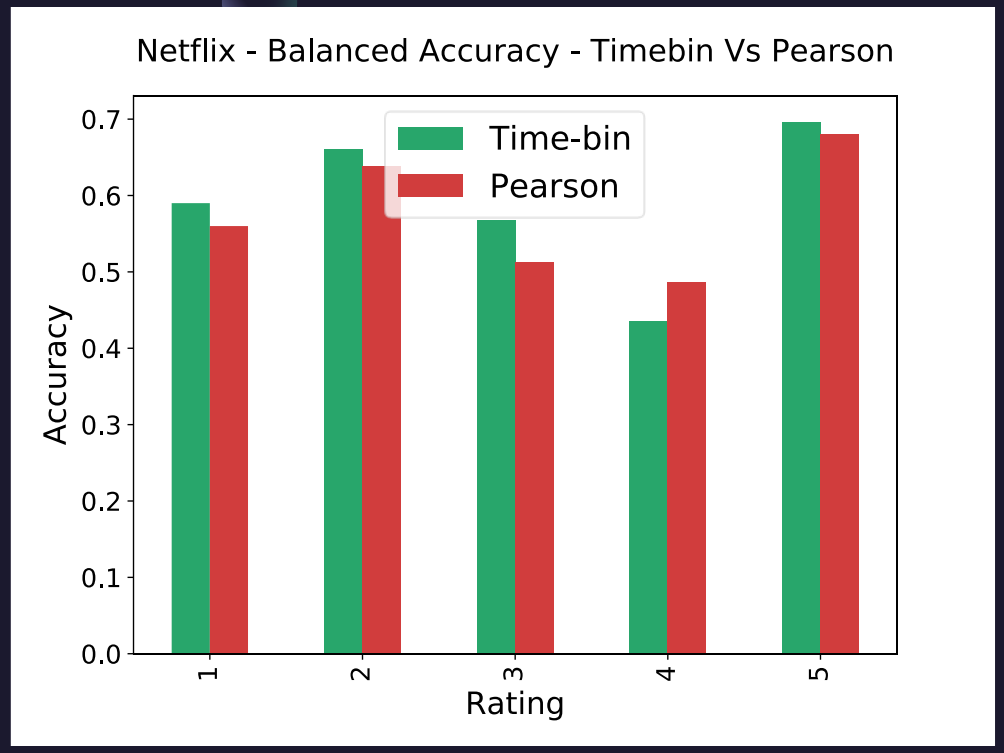

Movielens - F1 Score - Timebin Vs Pearson
Time-bin
Pearson
F1
0.0
0.1
0.2
0.3
0.4
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Rating Class

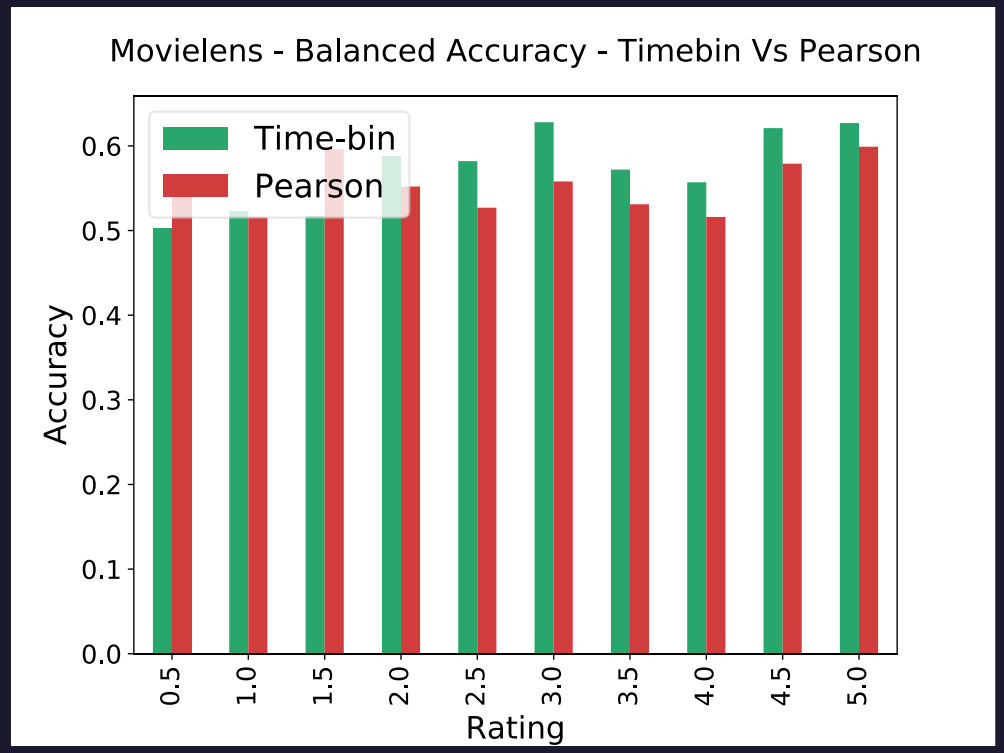
Accuracy


Netflix - Accuracy - Timebin Vs Pearson
Time-bin
Pearson
Accuracy
0.0
0.2
0.4
0.6
0.8
1
2
3
4
5
Rating

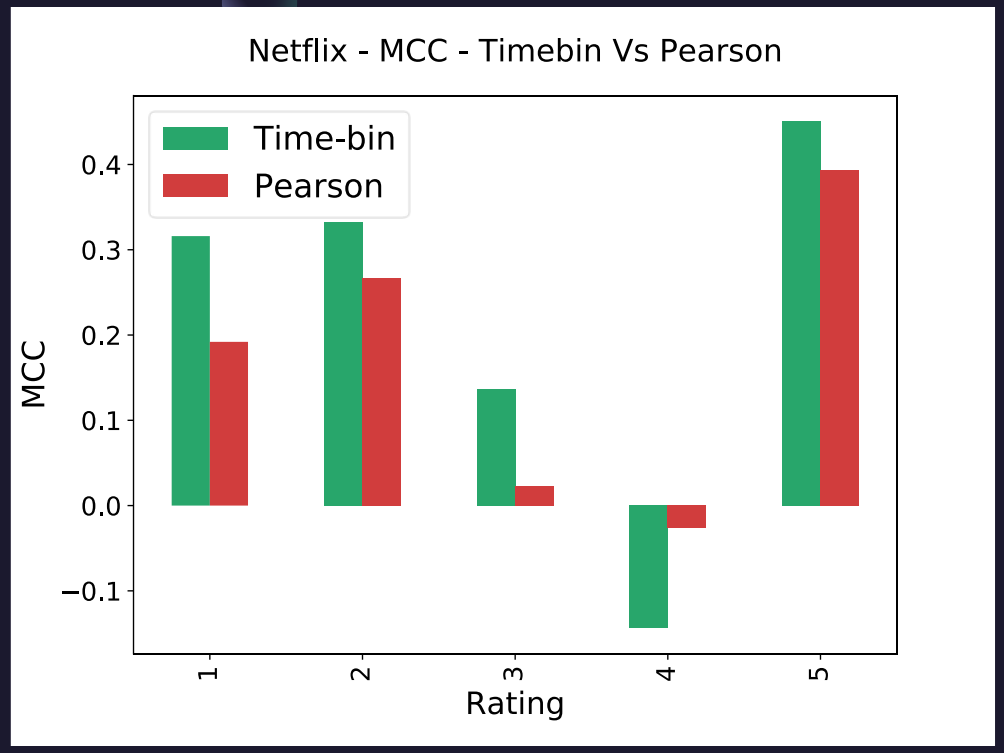

Movielens - Accuracy - Timebin Vs Pearson
Time-bin
Pearson
Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Rating

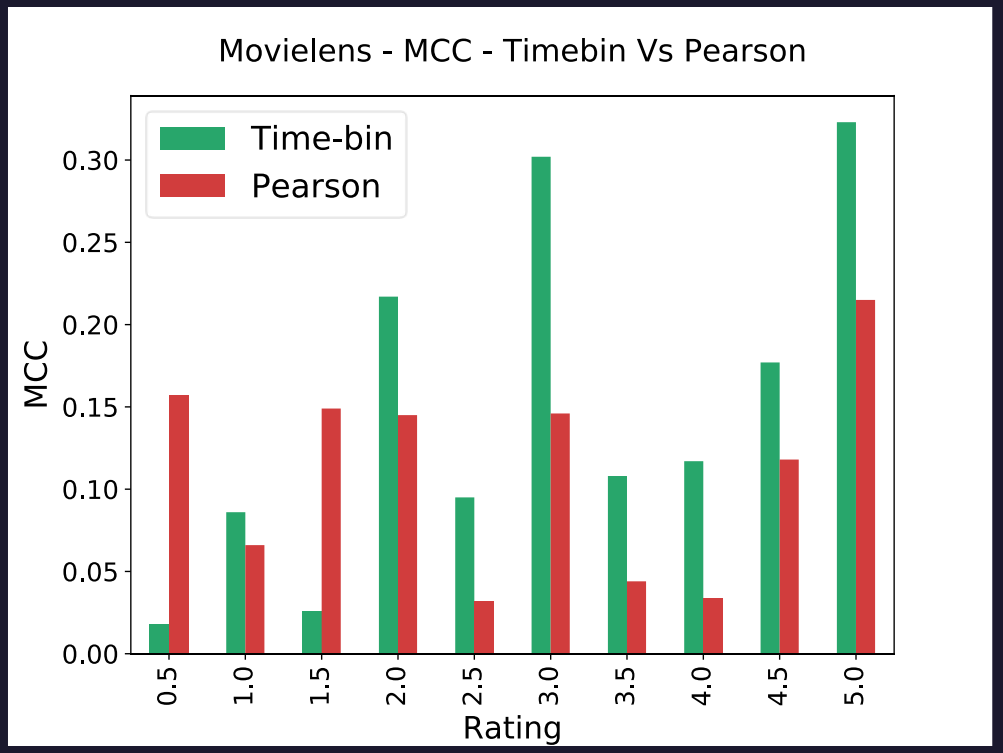
Balanced Accuracy



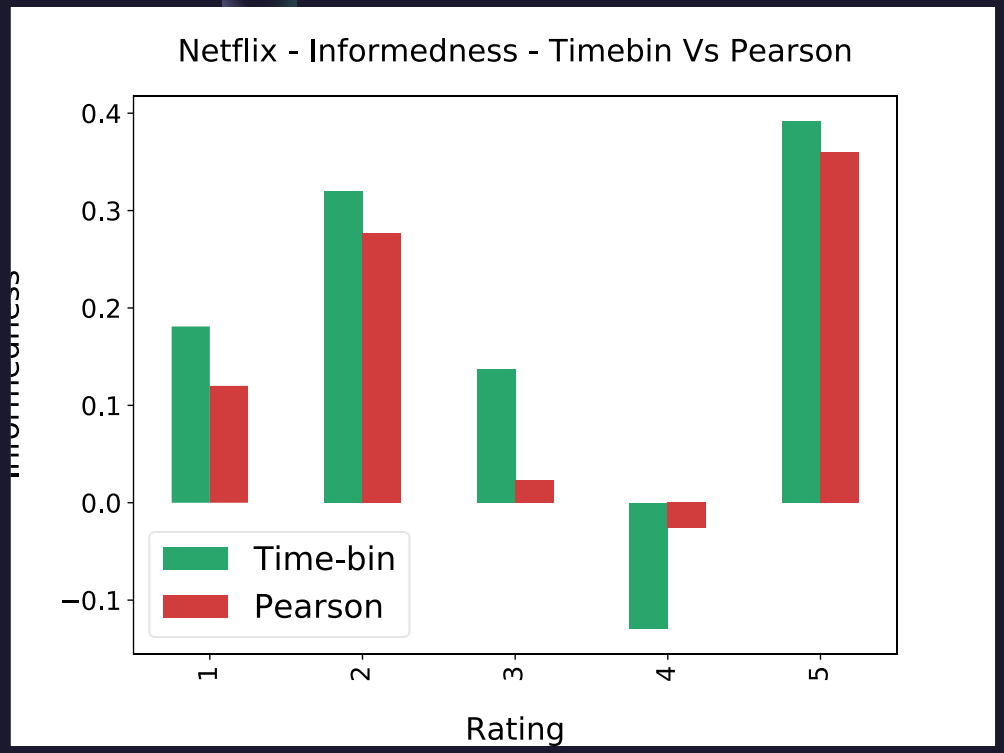


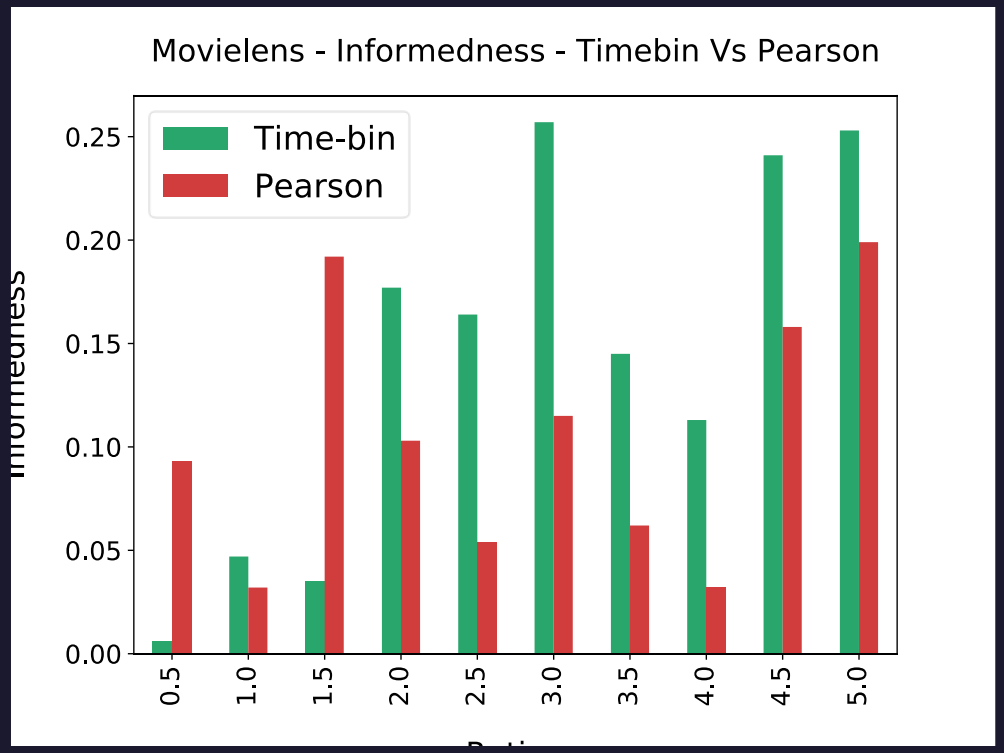
Matthews Correlation Coefficient



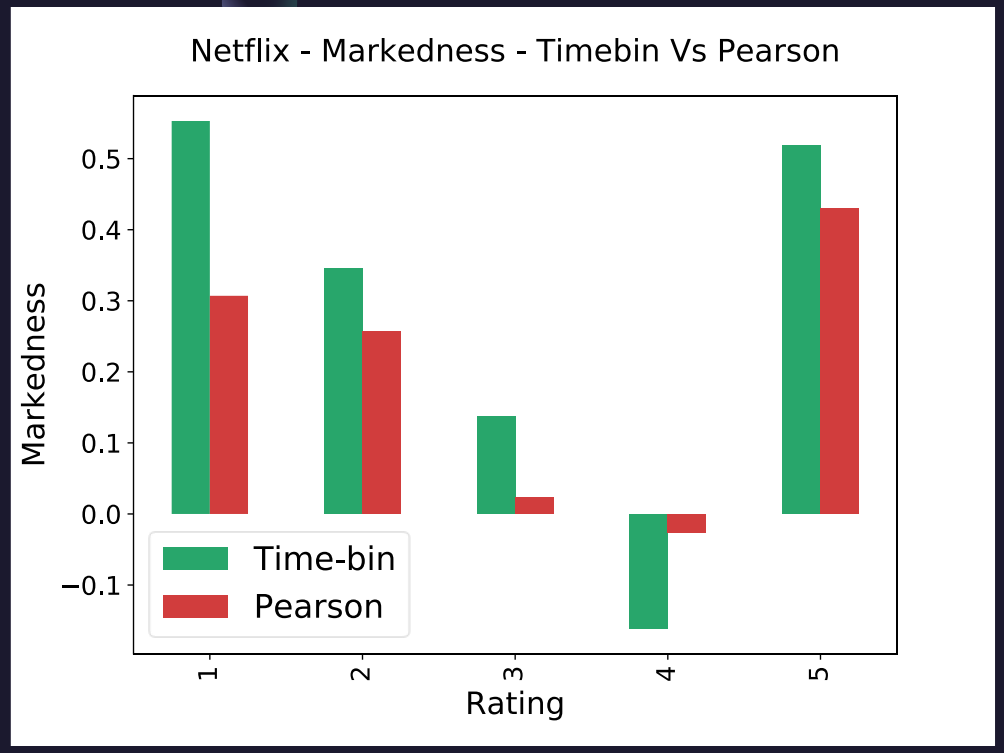


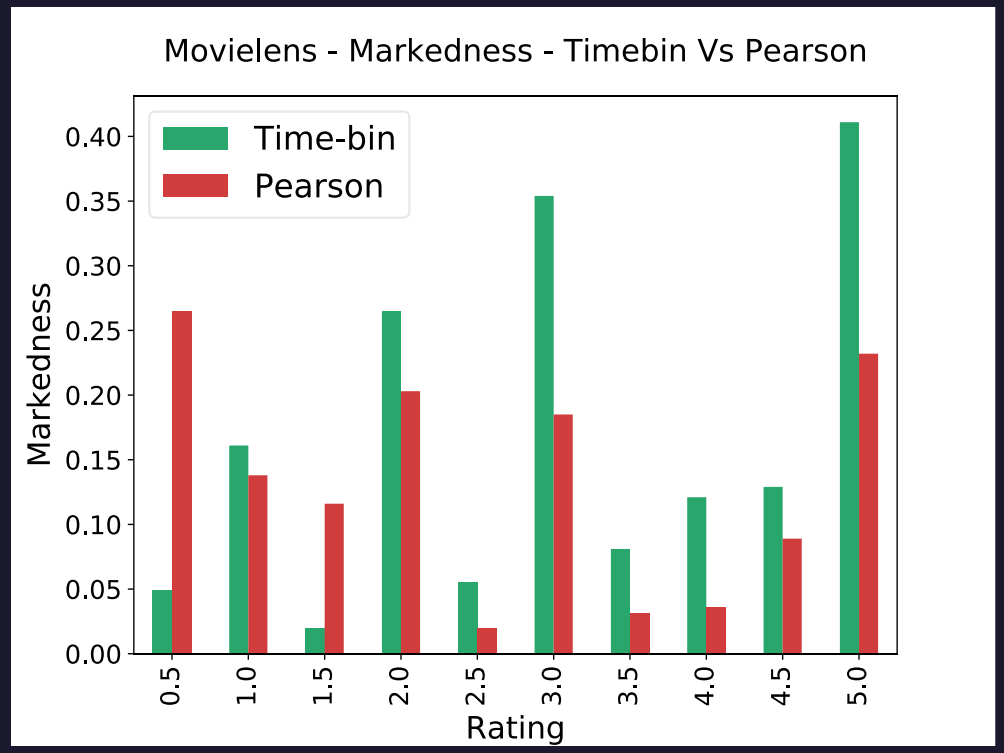
Informedness



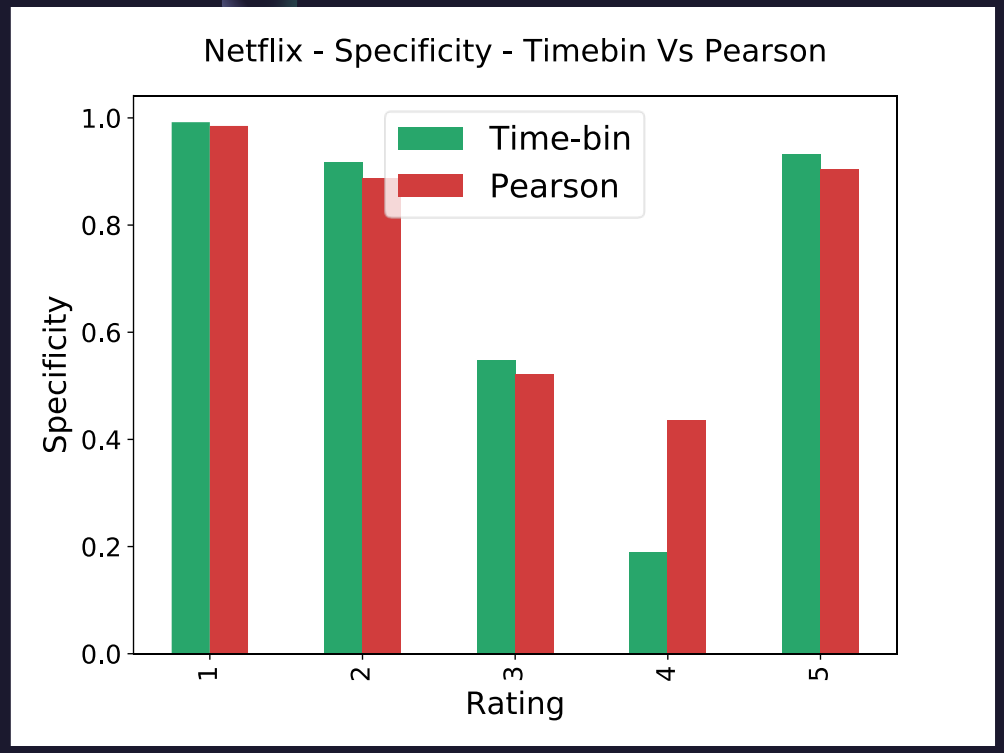


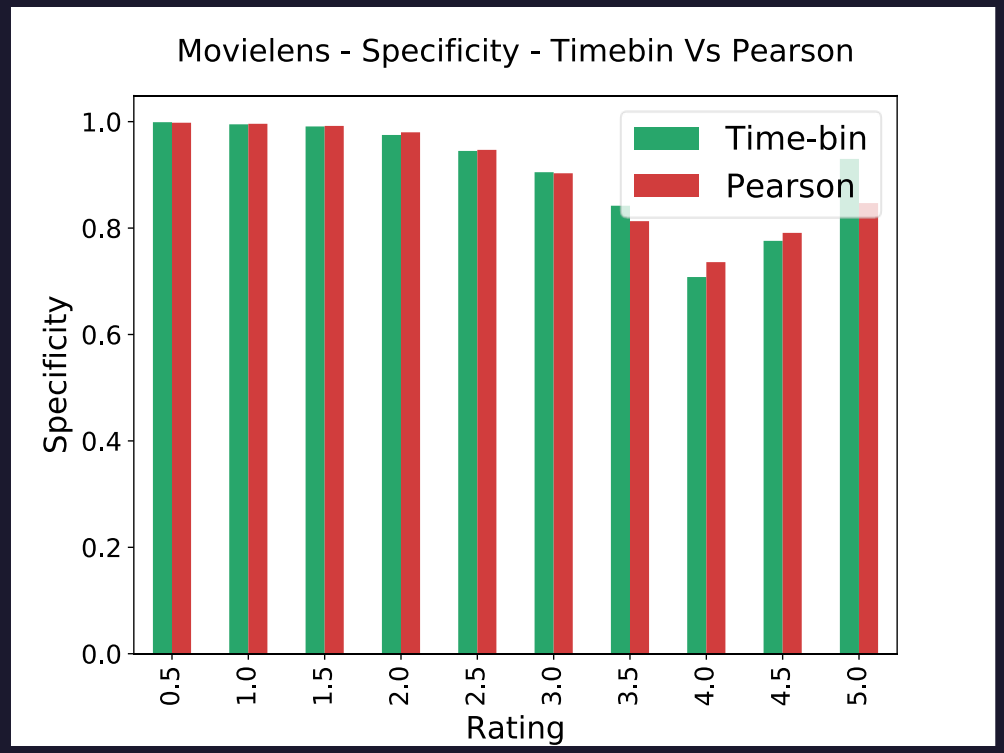
Markedness


Netflix - Markedness - Timebin Vs Pearson
Markedness
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
Time-bin
Pearson
1
2
3
4
5
Rating


Movielens - Markedness - Timebin Vs Pearson
Markedness
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Time-bin
Pearson
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Rating

Specificity





Conclusion

- We present Time-bin Based Neighborhood.
- Importance of temporal drifts.
- Space for future work on temporal drifts.

References

Our Temporal Drift Github Repository

One Against All Approach