Authors: Ryan Nett
Patrick Farrell

Collaborative Filtering and Recommender Systems

Methods

We implemented a variety of methods with different options that can all be used in conjunction with each other. All methods are implemented as both user and item-based with the option to include only the K nearest neighbors. For aggregation techniques we implemented both weighted sum and adjusted weighted sum; for similarity the choice is between cosine similarity and Pearson Correlation. This gave us a total of sixteen different method combinations, which are listed below with their corresponding IDs (the IDs are also provided when running the programs with no arguments).

Note that item-based means joke-based.

0 : item-based, no adjust, no knn, cosine similarity
1 : user-based, no adjust, no knn, cosine similarity
2 : item-based, adjusted, no knn, cosine similarity
3 : user-based, adjusted, no knn, cosine similarity
4 : item-based, no adjust, knn, cosine similarity
5 : user-based, no adjust, knn, cosine similarity
6 : item-based, adjusted, knn, cosine similarity
7 : user-based, adjusted, knn, cosine similarity
8 : item-based, no adjust, no knn, pearson similarity
9 : user-based, no adjust, no knn, pearson similarity
10 : item-based, adjusted, no knn, pearson similarity
11 : user-based, adjusted, no knn, pearson similarity
12 : item-based, no adjust, knn, pearson similarity
13 : user-based, no adjust, knn, pearson similarity
14 : item-based, adjusted, knn, pearson similarity
15 : user-based, adjusted, knn, pearson similarity

It is a bit overkill, but we implemented the methods as independent parameters (e.g. whether you use knn or not has no bearing on whether you adjust or not), which gave us all of these methods with minimal effort.

Research Question

When determining which method was “best”, we decided to look primarily at the number of jokes that were recommended by our system that were also rated highly (>=5) by the user. The reason we chose this was because in a recommender system we believe that the most important thing is that the recommendations are highly accurate, e.g. given a recommendation it is likely that the user will actually rate that item highly, not necessarily that it recommends a ton of items (quality over quantity). However, raw number of correctly recommended jokes does not give the complete picture. By recommending every single joke, regardless of whether the user is likely to rate it highly or not, a high number can be easily achieved at the cost of abysmal precision. To this end, we took the precision of the method into consideration as well so that a method that resulted in five correct recommendations and five incorrect recommendations would not be considered equal to one that resulted in five correct recommendations but only made one incorrect recommendation.

The question we used to determine the optimal method was therefore:

Which method has the highest accuracy with at least 75% precision?

Methodology

For testing we used 100 randomly generated (user ID, jokeID) pairs where the user had rated that joke with a single iteration for each of the methods listed above and input the resulting precision, recall, f1-measure and accuracies into a table so as to easily compare them.

Results

Below is the tabulated results of our testing for each of the methods. The program output (without the individual ratings) is given as well. Because we tested with only one iteration (repeats = 1), the standard deviation of the mean average errors is 0 for all program output because only one MAE was calculated.

Method	Precision	Recall	F1-Measure	Accuracy
0	0.0	0.0	0.0	0.79
1	0.0	0.0	0.0	0.77
2	0.60	0.16	0.25	0.82
3	0.45	0.19	0.27	0.73
4	0.78	0.23	0.36	0.75
5	0.5	0.29	0.37	0.76
6	0.57	0.18	0.28	0.79
7	0.83	0.17	0.28	0.74
8	0.66	0.09	0.16	0.79
9	0.0	0.0	0.0	0.75
10	0.66	0.1379	0.228	0.73
11	**0.88**	**0.33**	**0.48**	**0.83**
12	1.0	0.17	0.29	0.81
13	0.5	0.125	0.2	0.76
14	0.83	0.2	0.32	0.79
15	0.75	0.28	0.409	0.74

Program Output

0: Item-based, no adjust, no knn, cosine similarity

```
Mean MAE: 3.332410243460065
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        0               21
Irrelevant      0               79
Precision : 0.0
Recall : 0.0
F1-Measure : 0.0
Total Accuracy :  0.79
```

1: user-based, no adjust, no knn, cosine similarity

```
Mean MAE: 3.8861034582814553
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        0               23
Irrelevant      0               77
Precision : 0.0
Recall : 0.0
F1-Measure : 0.0
Total Accuracy :  0.77
```

2: Item-based, adjusted, no knn, cosine similarity

```
Mean MAE: 4.0290401408592595
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        3               16
Irrelevant      2               79
Precision : 0.6
Recall : 0.15789473684210525
F1-Measure : 0.25
Total Accuracy :  0.82
```

3: User-based, adjusted, no knn, cosine similarity

```
Mean MAE: 3.720208450516147
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        5               21
Irrelevant      6               68
Precision : 0.45454545454545453
Recall : 0.19230769230769232
F1-Measure : 0.27027027027027023
Total Accuracy :  0.73
```

4: Item-based, no adjust, knn, cosine similarity

```
Mean MAE: 3.7158276883933117
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        7               23
Irrelevant      2               68
Precision : 0.7777777777777778
Recall : 0.23333333333333334
F1-Measure : 0.35897435897435903
Total Accuracy :  0.75
```

5: User-based, no adjust, knn, cosine similarity

```
Mean MAE: 3.247220407678317
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        7               17
Irrelevant      7               69
Precision : 0.5
Recall : 0.2916666666666667
F1-Measure : 0.3684210526315789
Total Accuracy :  0.76
```

6: Item-based, adjusted, knn, cosine similarity

```
Mean MAE: 3.529542860400992
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        4               18
Irrelevant      3               75
Precision : 0.5714285714285714
Recall : 0.18181818181818182
F1-Measure : 0.27586206896551724
Total Accuracy :  0.79
```

7: User-based, adjusted, knn, cosine similarity

```
Mean MAE: 3.251666449263123
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        5               25
Irrelevant      1               69
Precision : 0.8333333333333334
Recall : 0.16666666666666666
F1-Measure : 0.2777777777777778
Total Accuracy :  0.74
```

8: Item-based, no adjustment, no knn, pearson similarity:

```
Mean MAE: 3.817632006882966
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        2               20
Irrelevant      1               77
Precision : 0.6666666666666666
```

```
Recall : 0.09090909090909091
F1-Measure : 0.16
Total Accuracy :  0.79
```

9 : User-based, no adjust, no knn, pearson similarity

```
Mean MAE: 4.334783610589833
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        0               25
Irrelevant      0               75
Precision : 0.0
Recall : 0.0
F1-Measure : 0.0
Total Accuracy :  0.75
```

10 : Item-based, adjusted, no knn, pearson similarity

```
Mean MAE: 3.3641937381146643
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        4               25
Irrelevant      2               69
Precision : 0.6666666666666666
Recall : 0.13793103448275862
F1-Measure : 0.2285714285714286
Total Accuracy :  0.73
```

11 : User-based, adjusted, no knn, pearson similarity

```
Mean MAE: 3.2912823453576276
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        8               16
Irrelevant      1               75
Precision : 0.8888888888888888
Recall : 0.3333333333333333
F1-Measure : 0.48484848484848486
Total Accuracy :  0.83
```

12 : Item-based, no adjust, knn, pearson similarity

```
Mean MAE: 3.3027969797242345
Stddev MAE: 0.0
```

```
                Recommended     Not Recommended
Relevant        4               19
Irrelevant      0               77
Precision : 1.0
Recall : 0.17391304347826086
F1-Measure : 0.29629629629629634
Total Accuracy :  0.81
```

13 : User-based, no adjust, knn, pearson similarity

```
Mean MAE: 3.588393212562851
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        3               21
Irrelevant      3               73
Precision : 0.5
Recall : 0.125
F1-Measure : 0.2
Total Accuracy :  0.76
```

14 : Item-based, adjusted, knn, pearson similarity

```
Mean MAE: 3.2418043664806566
Stddev MAE: 0.0
                Recommended     Not Recommended
Relevant        5               20
Irrelevant      1               74
Precision : 0.8333333333333334
Recall : 0.2
F1-Measure : 0.3225806451612903
Total Accuracy :  0.79
```

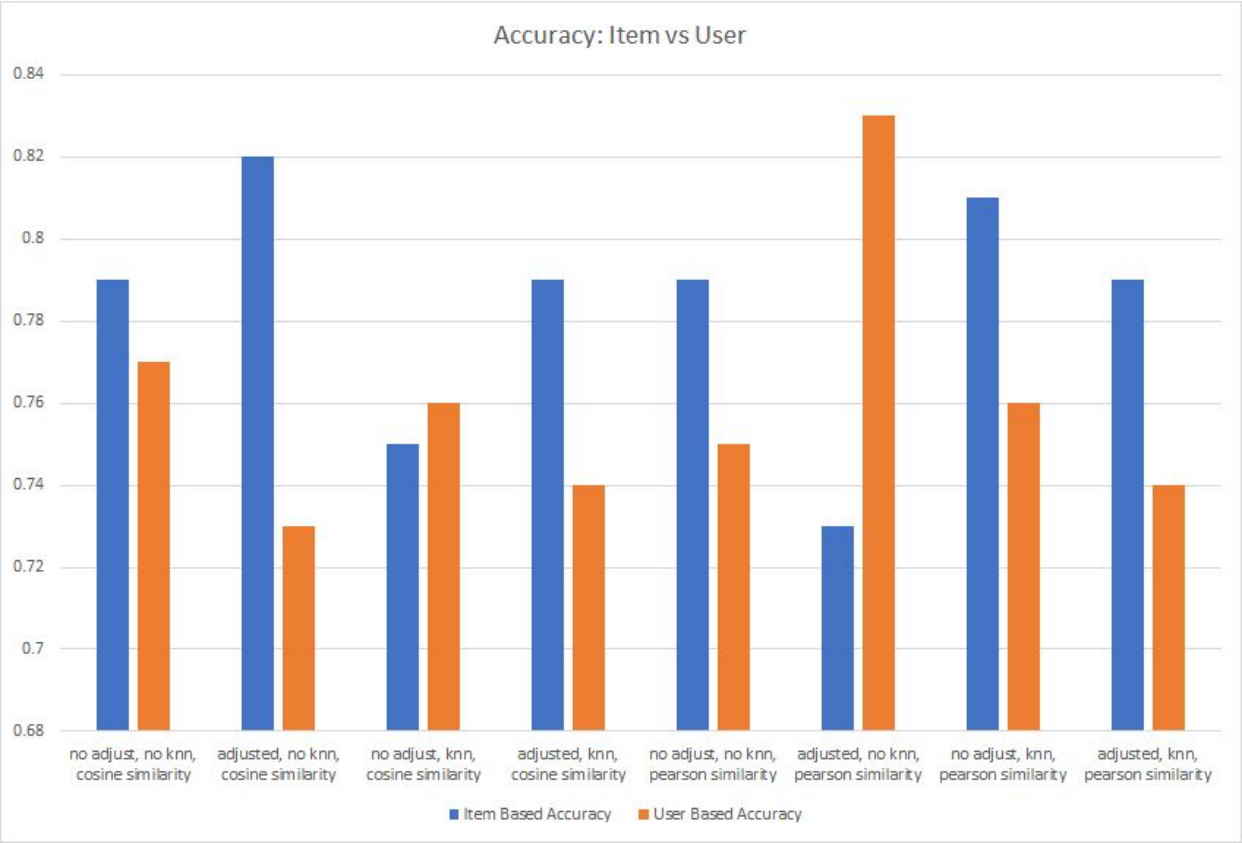
15 : User-based, adjusted, knn, pearson similarity

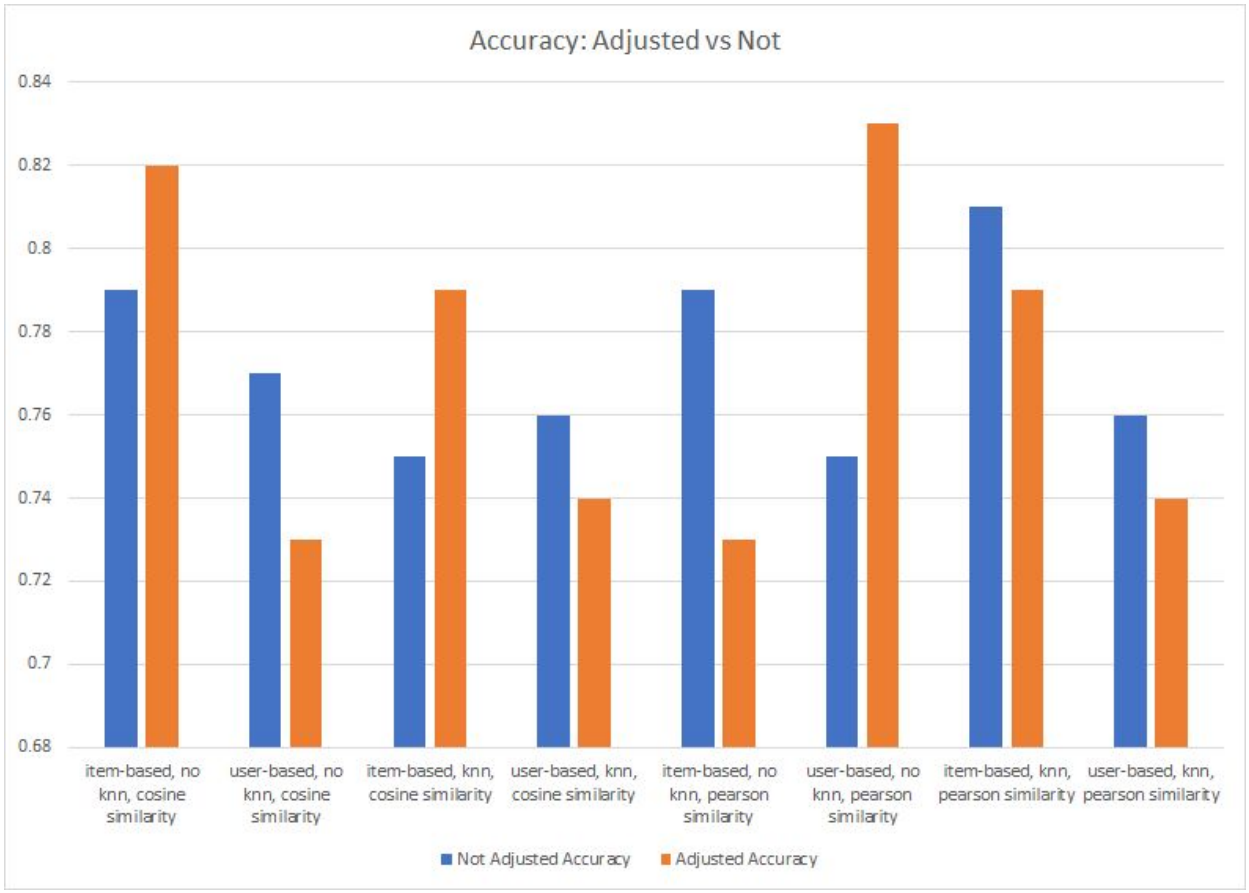
```
Mean MAE: 3.2507446670350704
Stddev MAE: 0.0
                Recommended    Not Recommended
Relevant        9              23
Irrelevant      3              65
Precision : 0.75
Recall : 0.28125
F1-Measure : 0.4090909090909091
Total Accuracy :  0.74
```

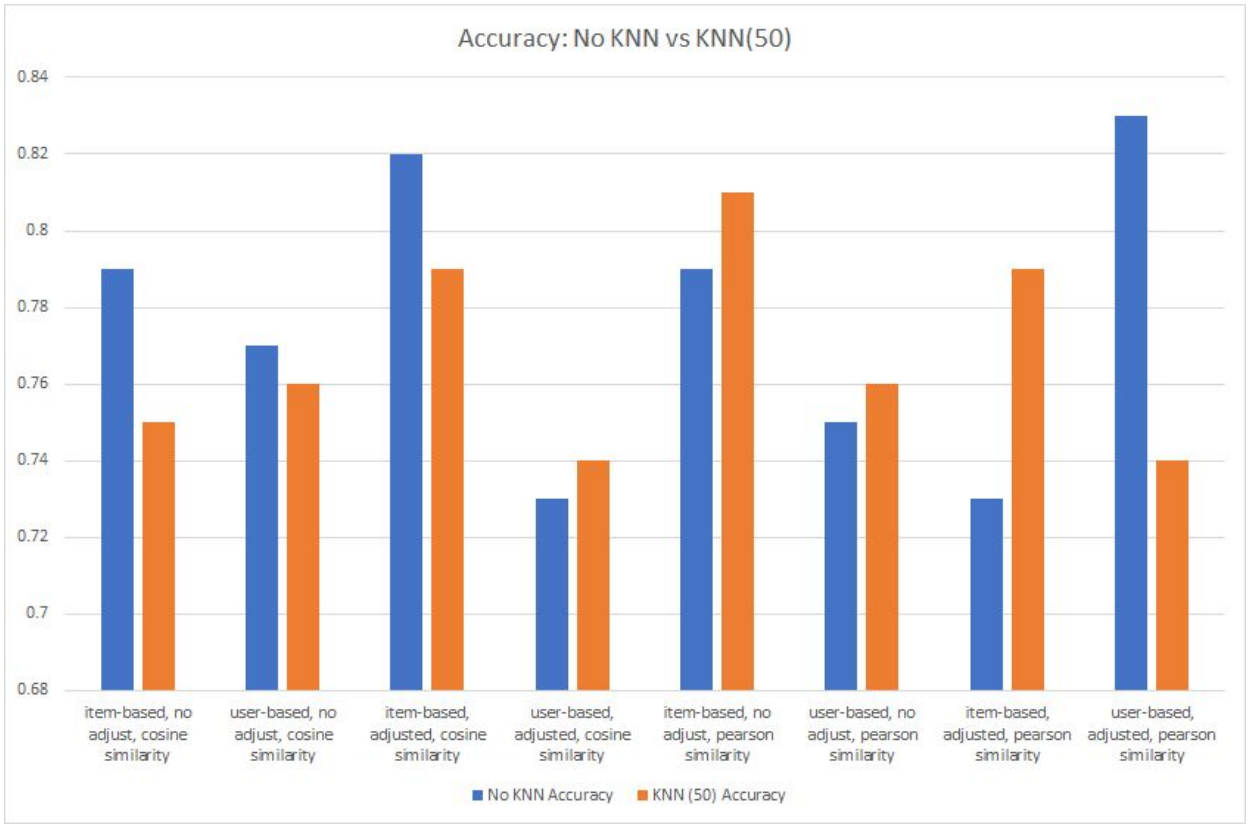
KNN testing for method 12

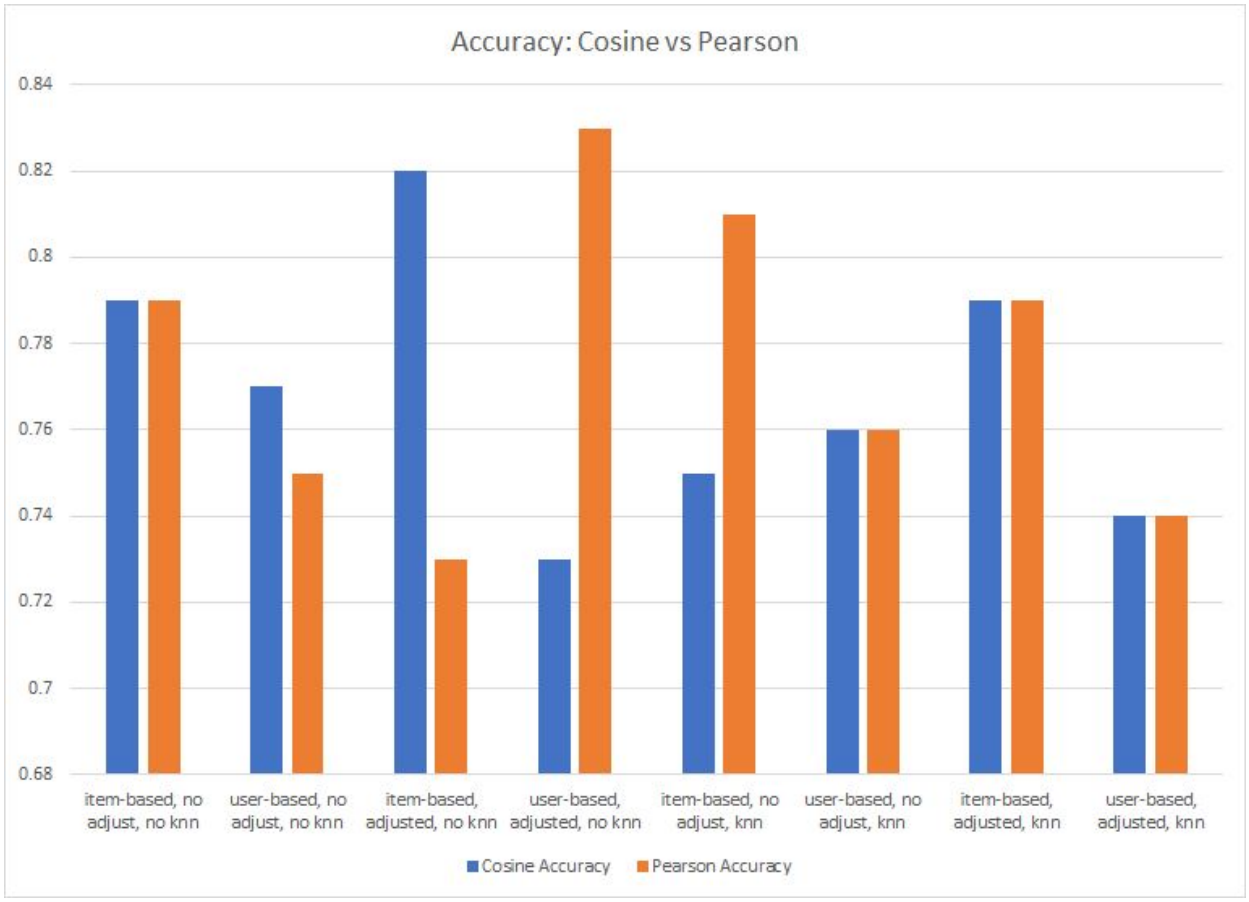
K	Precision	Recall	F1-Measure	Accuracy
10	*0.88*	*0.47*	*0.62*	*0.8*
20	0.88	0.25	0.39	0.78
30	1.0	0.08	0.15	0.78
40	1.0	0.09	0.16	0.79
50	1.0	0.17	0.29	0.81
60	1.0	0.11	0.20	0.84
70	0.67	0.08	0.14	0.75
80	1.0	0.07	0.13	0.74
90	0.6	0.11	0.19	0.74
100	1.0	0.05	0.10	0.62

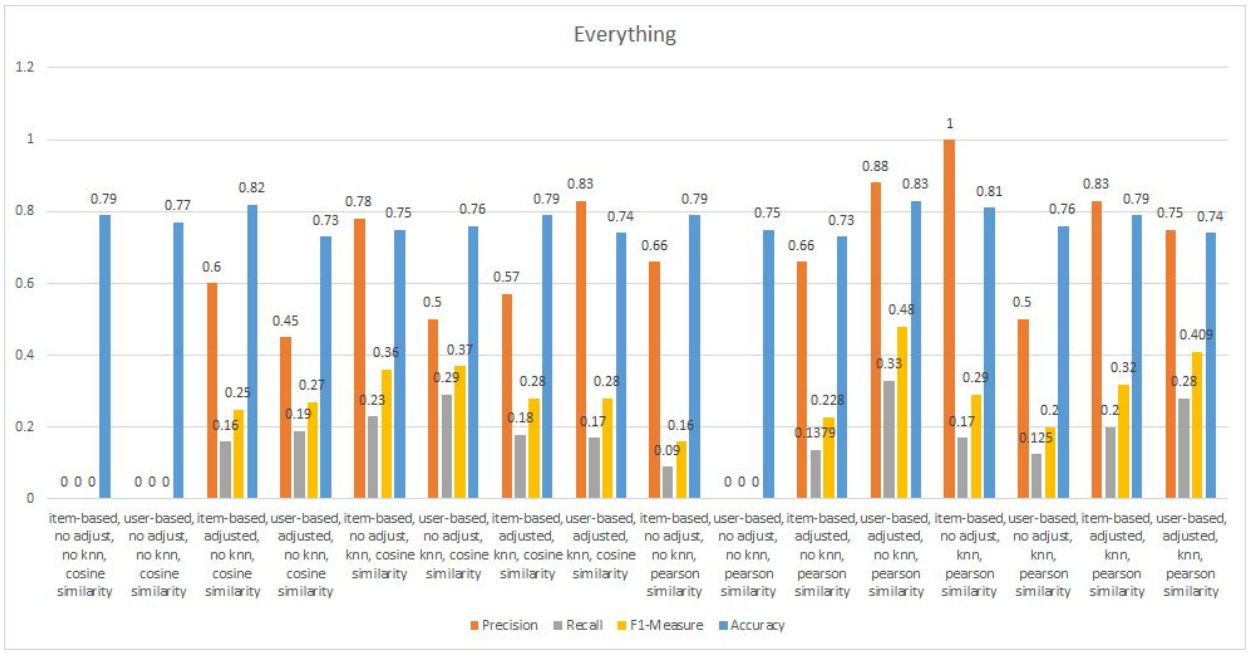
Conclusions

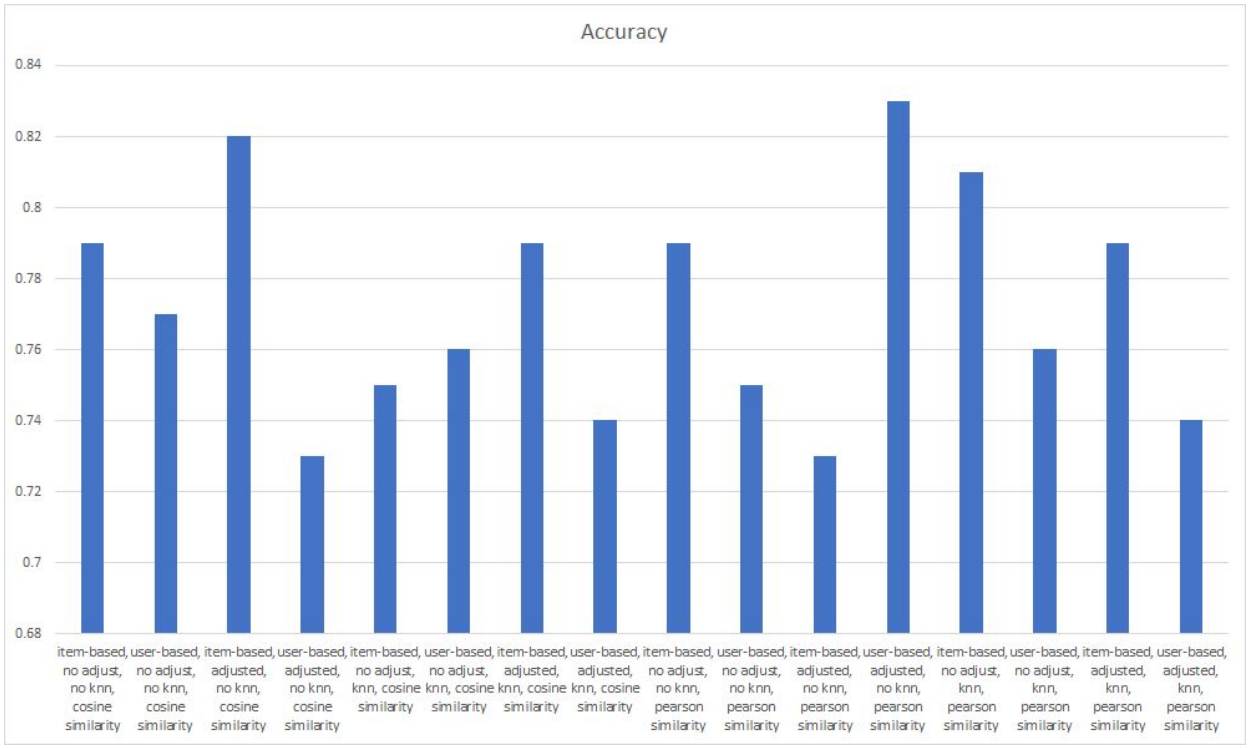


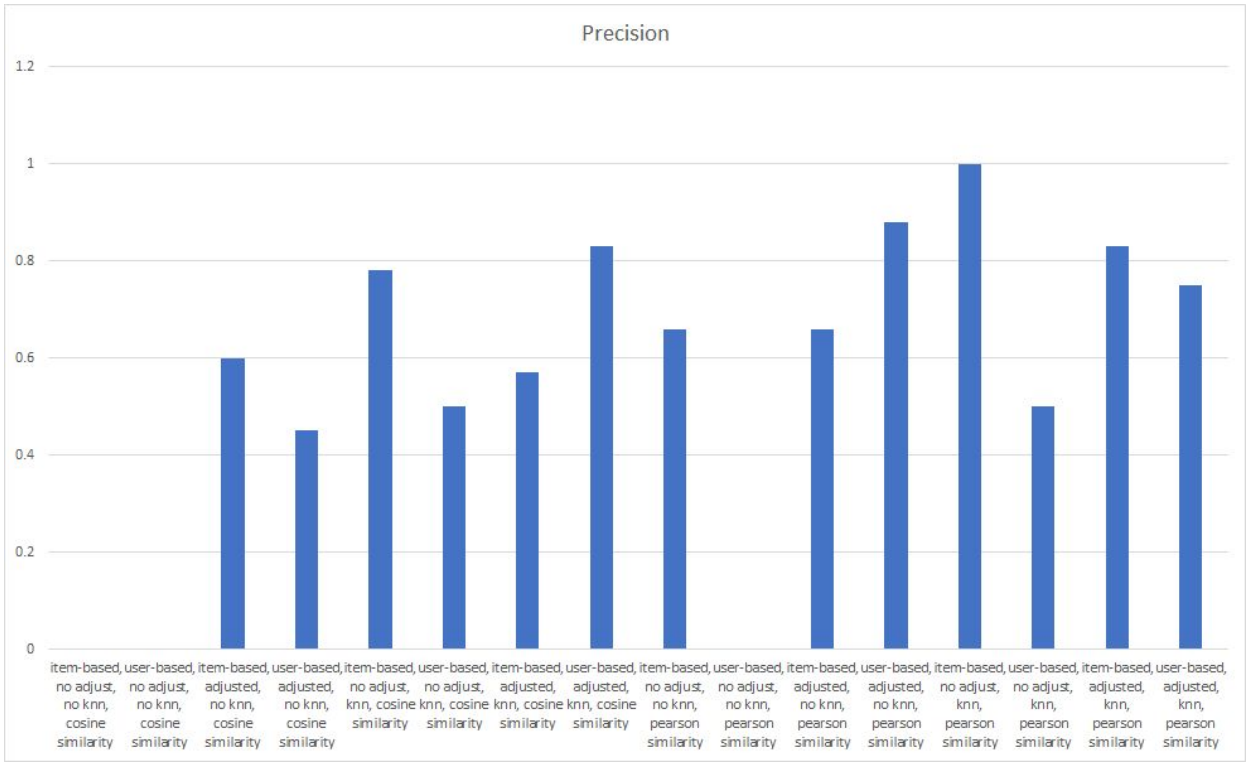

Accuracy: Adjusted vs Not
0.84
0.82
0.8
0.78
0.76
0.74
0.72
0.7
0.68
item-based, no knn, cosine similarity
user-based, no knn, cosine similarity
item-based, knn, cosine similarity
user-based, knn, cosine similarity
item-based, no knn, pearson similarity
user-based, no knn, pearson similarity
item-based, knn, pearson similarity
user-based, knn, pearson similarity
Not Adjusted Accuracy
Adjusted Accuracy

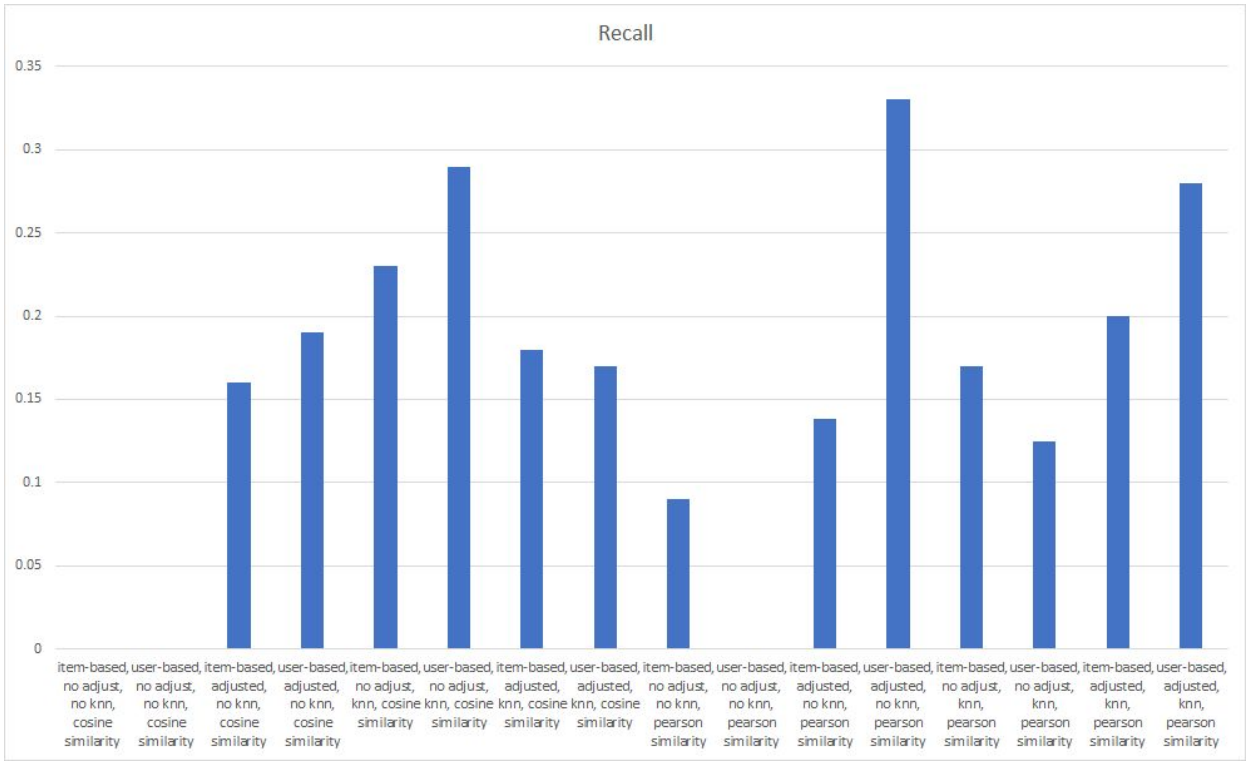

Accuracy: No KNN vs KNN(50)
0.84
0.82
0.8
0.78
0.76
0.74
0.72
0.7
0.68
item-based, no adjust, cosine similarity
user-based, no adjust, cosine similarity
item-based, adjusted, cosine similarity
user-based, adjusted, cosine similarity
item-based, no adjust, pearson similarity
user-based, no adjust, pearson similarity
item-based, adjusted, pearson similarity
user-based, adjusted, pearson similarity
No KNN Accuracy
KNN (50) Accuracy

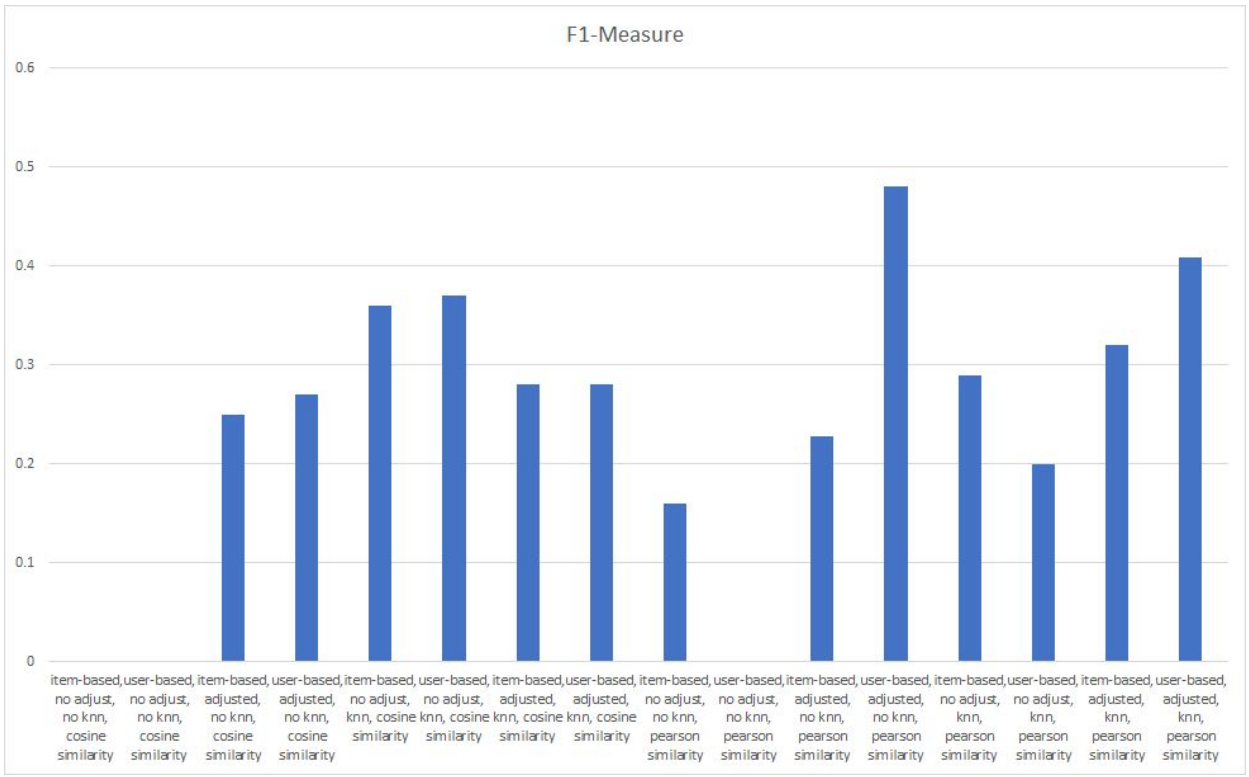

Accuracy: Cosine vs Pearson
0.84
0.82
0.8
0.78
0.76
0.74
0.72
0.7
0.68
item-based, no adjust, no knn
user-based, no adjust, no knn
item-based, adjusted, no knn
user-based, adjusted, no knn
item-based, no adjust, knn
user-based, no adjust, knn
item-based, adjusted, knn
user-based, adjusted, knn
Cosine Accuracy
Pearson Accuracy

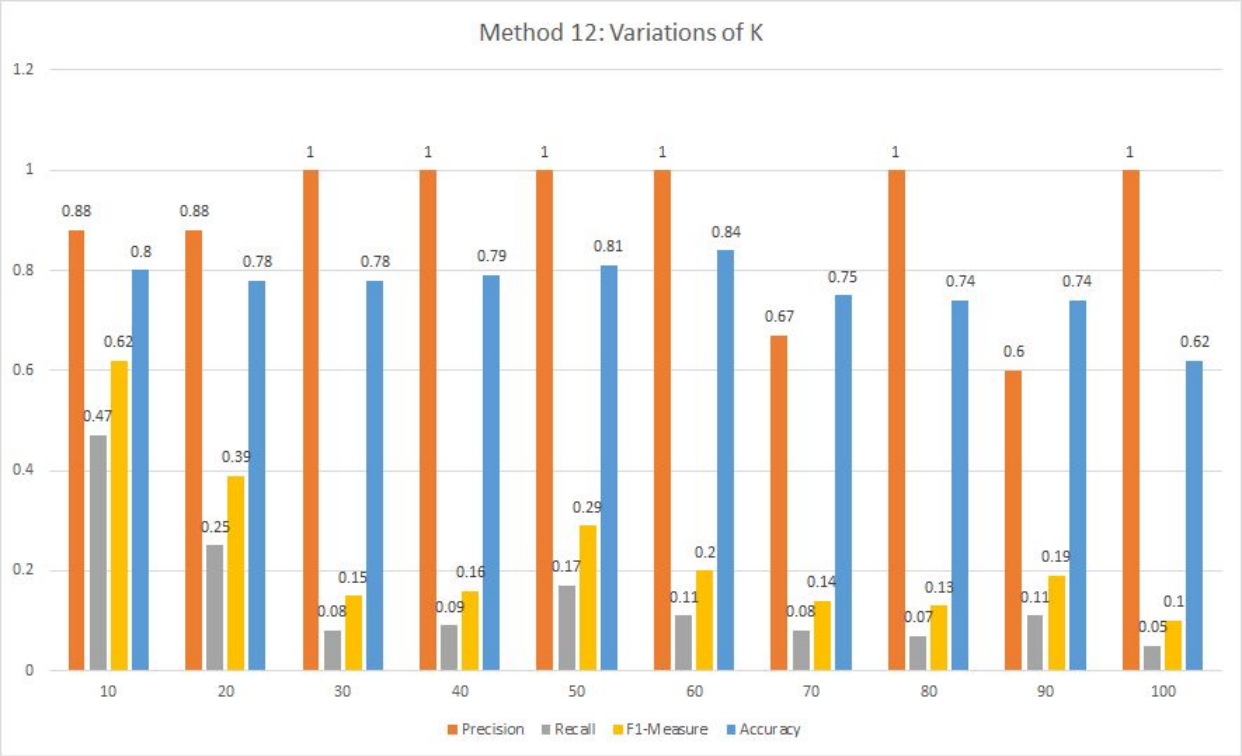

Everything
1.2
1
0.8
0.6
0.4
0.2
0
0 0 0
0.79
0 0 0
0.77
0.6
0.16
0.25
0.82
0.45
0.19
0.27
0.73
0.78
0.23
0.36
0.75
0.5
0.29
0.37
0.76
0.57
0.18
0.28
0.79
0.83
0.17
0.28
0.74
0.66
0.09
0.16
0.79
0 0 0
0.75
0.66
0.1379
0.228
0.73
0.88
0.33
0.48
0.83
1
0.17
0.29
0.81
0.5
0.125
0.2
0.76
0.83
0.2
0.32
0.79
0.75
0.28
0.409
0.74
item-based, no adjust, no knn, cosine similarity
user-based, no adjust, no knn, cosine similarity
item-based, adjusted, no knn, cosine similarity
user-based, adjusted, no knn, cosine similarity
item-based, no adjust, knn, cosine similarity
user-based, no adjust, knn, cosine similarity
item-based, adjusted, knn, cosine similarity
user-based, adjusted, knn, cosine similarity
item-based, no adjust, no knn, pearson similarity
user-based, no adjust, no knn, pearson similarity
item-based, adjusted, no knn, pearson similarity
user-based, adjusted, no knn, pearson similarity
item-based, no adjust, knn, pearson similarity
user-based, no adjust, knn, pearson similarity
item-based, adjusted, knn, pearson similarity
user-based, adjusted, knn, pearson similarity
Precision
Recall
F1-Measure
Accuracy


Accuracy
0.84
0.82
0.8
0.78
0.76
0.74
0.72
0.7
0.68
item-based, no adjust, no knn, cosine similarity
user-based, no adjust, no knn, cosine similarity
item-based, adjusted, no knn, cosine similarity
user-based, adjusted, no knn, cosine similarity
item-based, no adjust, knn, cosine similarity
user-based, no adjust, knn, cosine similarity
item-based, adjusted, knn, cosine similarity
user-based, adjusted, knn, cosine similarity
item-based, no adjust, no knn, pearson similarity
user-based, no adjust, no knn, pearson similarity
item-based, adjusted, no knn, pearson similarity
user-based, adjusted, no knn, pearson similarity
item-based, no adjust, knn, pearson similarity
user-based, no adjust, knn, pearson similarity
item-based, adjusted, knn, pearson similarity
user-based, adjusted, knn, pearson similarity


Precision
1.2
1
0.8
0.6
0.4
0.2
0
item-based, no adjust, no knn, cosine similarity
user-based, no adjust, no knn, cosine similarity
item-based, adjusted, no knn, cosine similarity
user-based, adjusted, no knn, cosine similarity
item-based, no adjust, knn, cosine similarity
user-based, no adjust, knn, cosine similarity
item-based, adjusted, knn, cosine similarity
user-based, adjusted, knn, cosine similarity
item-based, no adjust, no knn, pearson similarity
user-based, no adjust, no knn, pearson similarity
item-based, adjusted, no knn, pearson similarity
user-based, adjusted, no knn, pearson similarity
item-based, no adjust, knn, pearson similarity
user-based, no adjust, knn, pearson similarity
item-based, adjusted, knn, pearson similarity
user-based, adjusted, knn, pearson similarity


Recall
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
item-based, no adjust, no knn, cosine similarity
user-based, no adjust, no knn, cosine similarity
item-based, adjusted, no knn, cosine similarity
user-based, adjusted, no knn, cosine similarity
item-based, no adjust, knn, cosine similarity
user-based, no adjust, knn, cosine similarity
item-based, adjusted, knn, cosine similarity
user-based, adjusted, knn, cosine similarity
item-based, no adjust, no knn, pearson similarity
user-based, no adjust, no knn, pearson similarity
item-based, adjusted, no knn, pearson similarity
user-based, adjusted, no knn, pearson similarity
item-based, no adjust, knn, pearson similarity
user-based, no adjust, knn, pearson similarity
item-based, adjusted, knn, pearson similarity
user-based, adjusted, knn, pearson similarity


F1-Measure
0.6
0.5
0.4
0.3
0.2
0.1
0
item-based, no adjust, no knn, cosine similarity
user-based, no adjust, no knn, cosine similarity
item-based, adjusted, no knn, cosine similarity
user-based, adjusted, no knn, cosine similarity
item-based, no adjust, knn, cosine similarity
user-based, no adjust, knn, cosine similarity
item-based, adjusted, knn, cosine similarity
user-based, adjusted, knn, cosine similarity
item-based, no adjust, no knn, pearson similarity
user-based, no adjust, no knn, pearson similarity
item-based, adjusted, no knn, pearson similarity
user-based, adjusted, no knn, pearson similarity
item-based, no adjust, knn, pearson similarity
user-based, no adjust, knn, pearson similarity
item-based, adjusted, knn, pearson similarity
user-based, adjusted, knn, pearson similarity



One of the problems with using so many methods is that it is hard to compare them all. We found that splitting the results up by one choice (e.g. adjusted or not adjusted) and the graphing the results helped. Unfortunately, these graphs are rather time consuming to make, so we only made them for accuracy.

We feel like there is a better way to visualize the data based on these 4 decisions, that would allow us to see that say, adjusted is better for item-based but worse for user-based, but we didn't know of any.

These graphs show that generally, item-based is better than user-based. Although for adjusted pearson with no knn, user is far better.

Adjusted vs not adjusted is a wash, it is better in some cases and worse in others.

KNN tended to be worse for cosine similarity, but better for pearson. Except for user-based, adjusted, where this was reversed.

There were quite a few where the accuracy was the same with cosine or pearson. However, there were two cases whee pearson preformed vastly better, and only one (and a small one) where cosine did.

The best method was method 11: user-based, adjusted, no knn pearson. It had the best accuracy, recall, and f-measure, and the second best precision. It achieved 83% accurate recommendations, with a MAE of 3.3.

However, this does not take into account the variability of k values.The best KNN method, by a significant margin, was 12: item based, no adjust, knn, pearson. We ran extra tests on different K values (this is the last section in **Results**).

The best K value turned out to be 10, with 80% accuracy. Even though 50 has a higher accuracy, it had much less recall. Since the accuracy was only 1% better, we decided that 10 was better. **This was not quite enough to beat method 11, so 11 (user-based, adjusted, no knn pearson) remains as our best method.**