



# Stream to Success

---

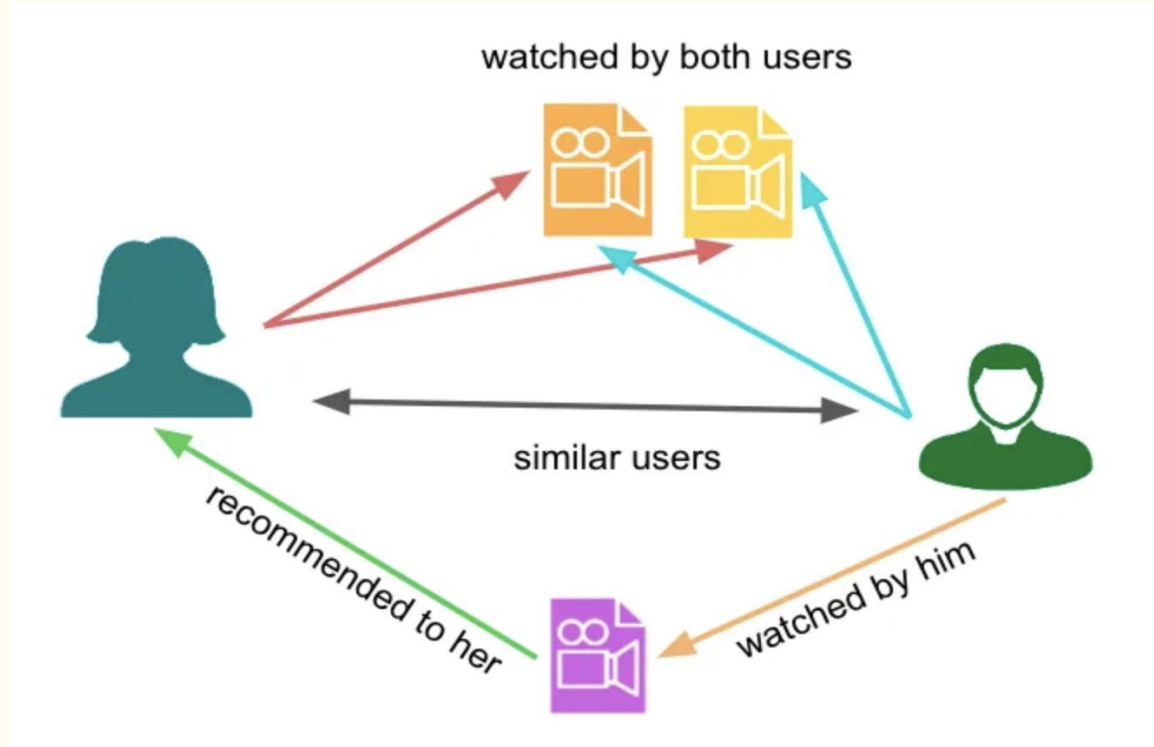
An Exploratory Approach to Personalized Recommendation Systems

# The Dataset

- Small MovieLens dataset
  - GroupLens research lab at the University of Minnesota
- ratings.csv
- movies.csv
- tags.csv
- links.csv



# Collaborative Filtering



# How we're evaluating our algorithms:

RMSE: Root Mean Squared Error

- Illustrates the difference between *predicted* ratings and *actual* ratings.
- The more each score approaches 0 (aka, *no* difference between predicted and actual ratings), the better.
- We are looking for a *low* RMSE score.

# ALS Algorithm

Sample user output: user 123

RMSE: 0.5705045822613408

On average, the ALS model's predicted ratings deviate from the actual ratings by approximately 0.57 stars.

movieId	title	prediction
51255	Hot Fuzz (2007)	5.304486
1356	Star Trek: First Contact (1996)	4.9631844
3949	Requiem for a Dream (2000)	4.785082
231	Dumb & Dumber (Dumb and Dumber) (1994)	4.777913
339	While You Were Sleeping (1995)	4.766075
253	Interview with the Vampire: The Vampire Chronicles (1994)	4.693458
736	Twister (1996)	4.6628013
2194	Untouchables, The (1987)	4.575135
364	Lion King, The (1994)	4.5734696
6	Heat (1995)	4.545358

# SVD Algorithm

Sample user output: User 300

RMSE: 0.4751

On average, the SVD model's predictions deviate from *actual* ratings by around 0.48 stars. This is an improvement on the ALS model's RMSE score.

1. Heat (1995) - Predicted Rating: 4.50
2. Seven (a.k.a. Se7en) (1995) - Predicted Rating: 4.47
3. Usual Suspects, The (1995) - Predicted Rating: 4.42
4. Twelve Monkeys (a.k.a. 12 Monkeys) (1995) - Predicted Rating: 4.39
5. Casino (1995) - Predicted Rating: 4.33
6. Crimson Tide (1995) - Predicted Rating: 4.32
7. Taxi Driver (1976) - Predicted Rating: 4.30
8. Braveheart (1995) - Predicted Rating: 4.25
9. Title Not Found - Predicted Rating: 4.21
10. Title Not Found - Predicted Rating: 4.21

# KNNWithMeans Algorithm

Sample user output: user 59

RMSE: 0.3822

On average, the KNN model's predictions deviate from *actual* ratings by around 0.38 stars. This is an improvement upon both the ALS and SVD models' scores.

```
Movie ID: 1199, Title: Brazil (1985), Predicted Rating: 4.873752033120196
Movie ID: 1732, Title: Big Lebowski, The (1998), Predicted Rating: 4.867298998651436
Movie ID: 318, Title: Shawshank Redemption, The (1994), Predicted Rating: 4.864628499770431
Movie ID: 2324, Title: Life Is Beautiful (La Vita è bella) (1997), Predicted Rating: 4.86230109542257
Movie ID: 1203, Title: 12 Angry Men (1957), Predicted Rating: 4.843049925195593
Movie ID: 8874, Title: Shaun of the Dead (2004), Predicted Rating: 4.839640816089479
Movie ID: 50, Title: Usual Suspects, The (1995), Predicted Rating: 4.8312855504684595
Movie ID: 7361, Title: Eternal Sunshine of the Spotless Mind (2004), Predicted Rating: 4.815897748388172
Movie ID: 5618, Title: Spirited Away (Sen to Chihiro no kamikakushi) (2001), Predicted Rating: 4.814827992589781
Movie ID: 3949, Title: Requiem for a Dream (2000), Predicted Rating: 4.79296328245507
```

# Tuned KNNWithMeans Algorithm

We tuned the KNNWithMeans algorithm and got an even lower RMSE.

Here we see a list of all models and their corresponding RMSE scores.

	<b>Model</b>	<b>RMSE</b>
<b>0</b>	ALS	0.570505
<b>1</b>	SVD	0.475072
<b>2</b>	KNN	0.382185
<b>3</b>	KNN2	0.318239



Continue to fine tune  
the *KNNWithMeans*  
model.

1. Scale up dataset.
2. Explore additional metrics, like NDCG or MAP.
3. Implement user feedback.
4. Explore a content-based or hybrid approach.

# Thank you. Let's Connect!

E-mail:

[mbirchhn@gmail.com](mailto:mbirchhn@gmail.com)

Github:

[www.github.com/madelinebirch](https://www.github.com/madelinebirch)

LinkedIn:

<https://www.linkedin.com/in/madeline-birch-164000b5/>

Questions?