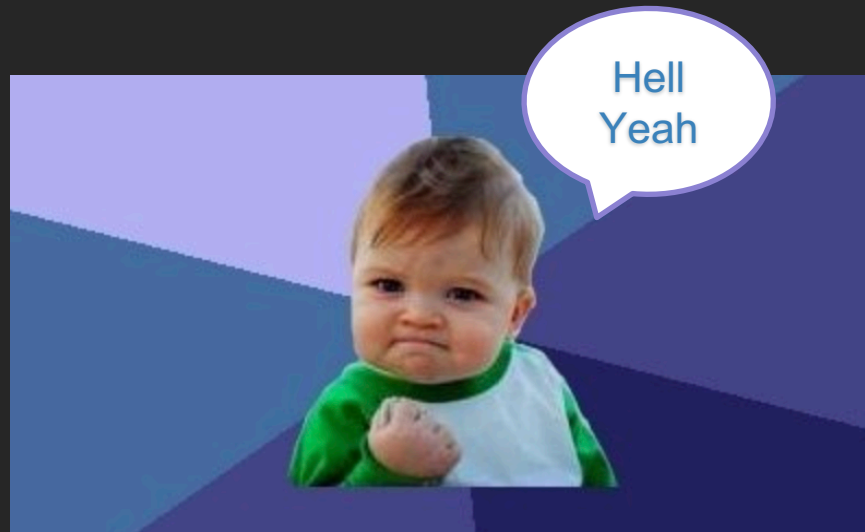# Need a Date?

Jenny Tang

Project Overview

Partnered with an online dating website to build a recommender system to maximize successful matches.

Raw Data

- 1 million user profiles from past 10 years
- 27 million actions in 2017
- Set threshold for 200,000 active users in 2017 with 10 million actions
- Due to the landscape of the data, focus on heterosexual relationships
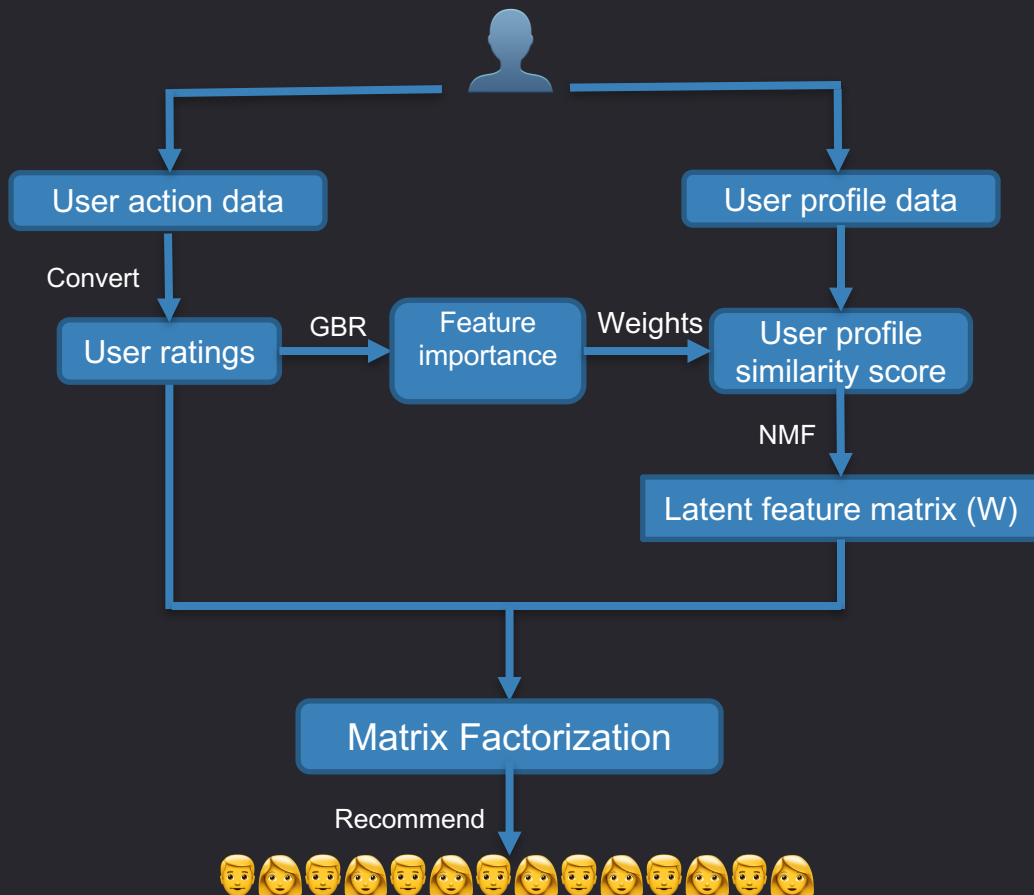
Two Types of Users

Existing Users

- Certain level of **profile** completion

- Lots of online **actions**
  visitations, bookmarks, messages…

- Threshold: # of actions >= 10

New Users / Inactive Users

- Certain level of **profile** completion

- Lack of online **actions**
  Haven't viewed or been viewed..

- Threshold: # of actions < 10

Model for Existing Users

- Convert actions (implicit) to ratings (explicit)
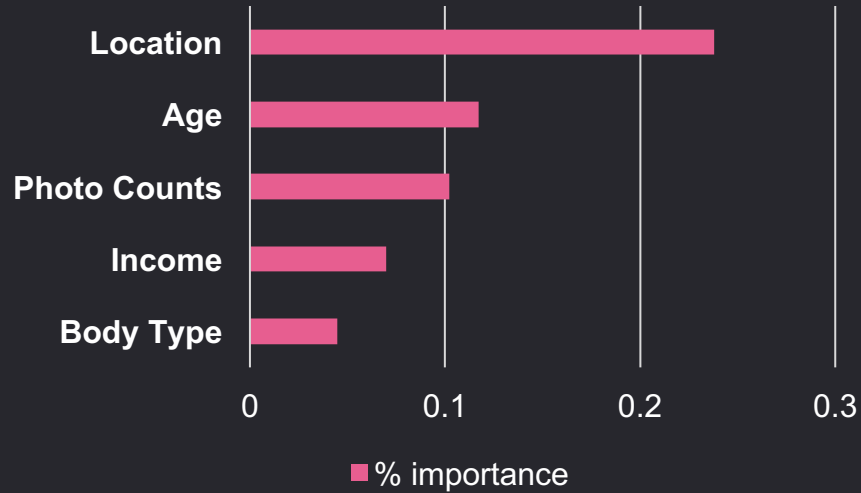
  Visitation -> 1
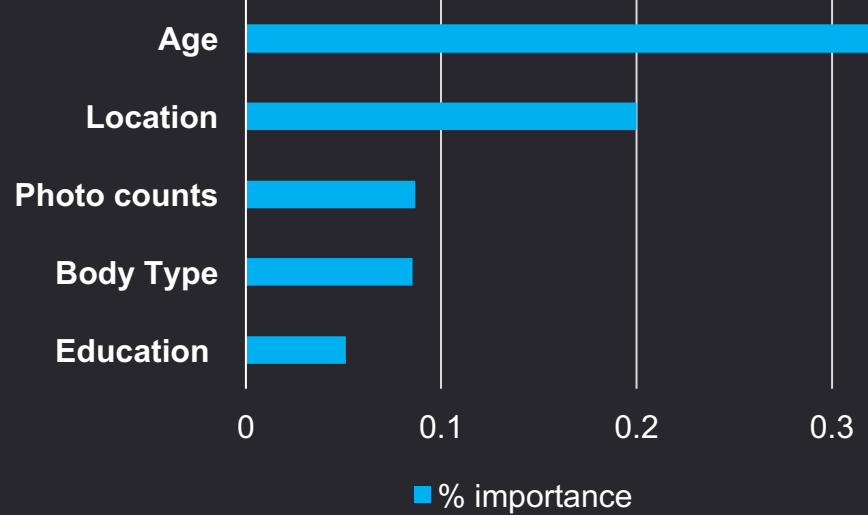 *Bookmark -> 2*
  Favorite -> 3
  Message -> 4

- Rating received (Popularity Score) used as target in Gradient boosting Regression

- Capture the weights of the profile features related to one's popularity score

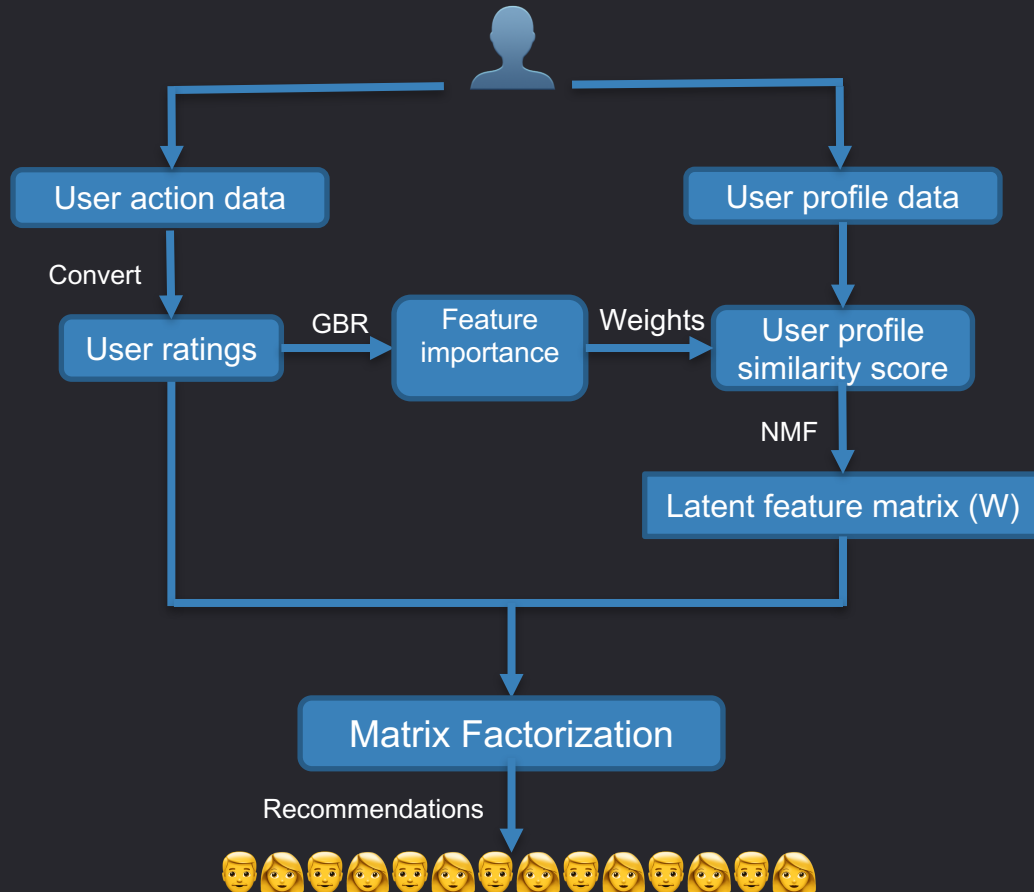# Top Five Features

For Women

	% importance
Location	
Age	
Photo Counts	
Income	
Body Type	

For Men

	% importance
Age	
Location	
Photo counts	
Body Type	
Education	

Model for Existing Users

- Convert actions (implicit) to ratings (explicit)

  Visitation -> 1
  Bookmark -> 2
  Favorite -> 3
  Message -> 4

- R received = Popularity Score used as target in Gradient boosting Regression

- Capture the weights of the profile features related to one's popularity score

New User Challenge

- New users have not received or given ratings (actions)
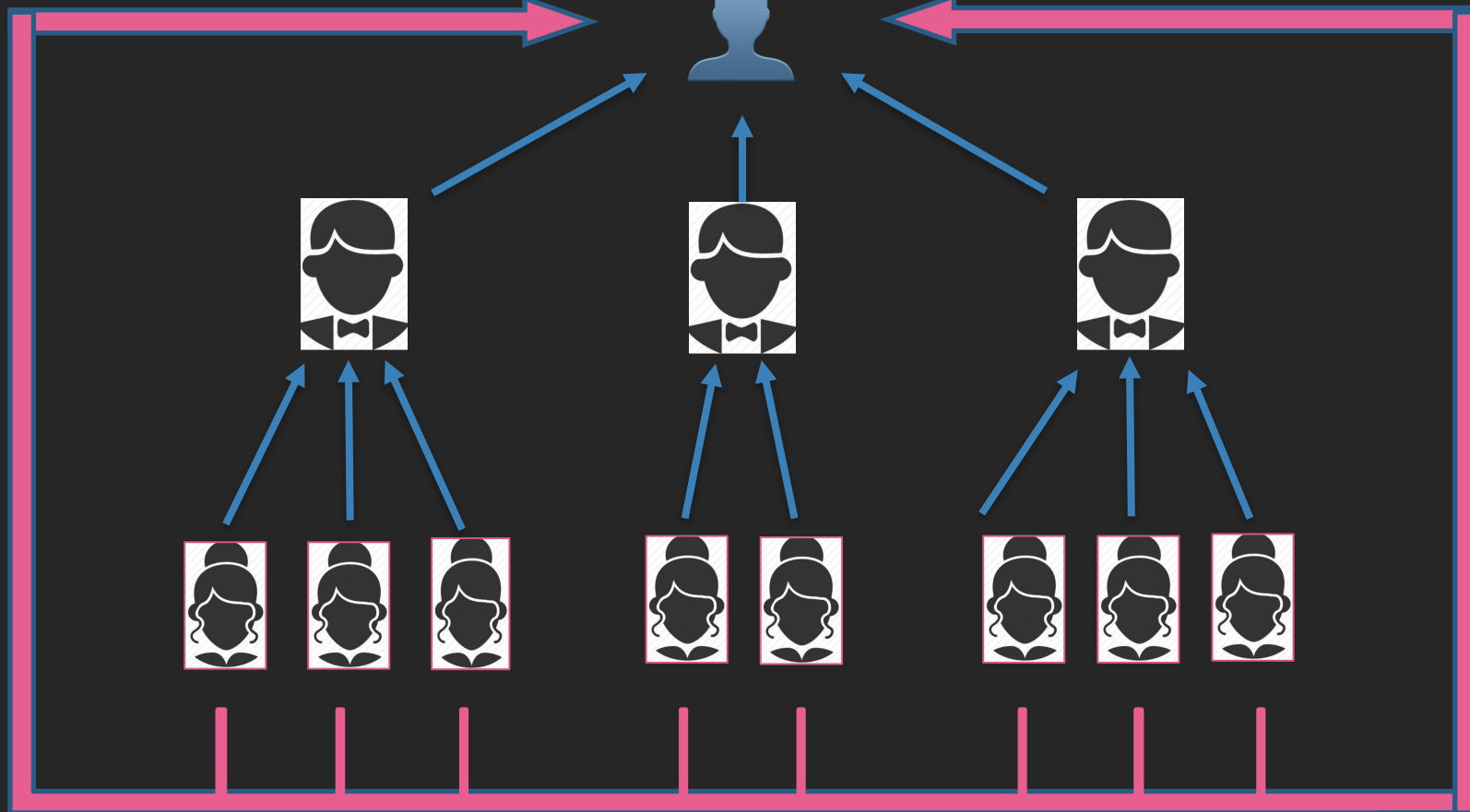
Solution

- Reverse match
- Search for who may be interested in this new user among existing users

Recommender for New Users

What happens after cold users are getting "hot"?

New users actions reach the thresholds

Existing users

Welcome to *Recommender System for Existing User*

Model Comparison

Baseline Model

- Distance + Popularity

- Not customized

- Overlook combinations of preferences

Model for Existing Users

✓ **Hybrid** recommender system

✓ User profile similarity as **Content Based Filtering**

✓ Historical user actions as **Collaborative Filtering**

Model for New Users

✓ Reverse Match

✓ There are always matches for new users! NOT always recommend popular people

✓ Increase the match rate

Evaluation

$$Quality = \frac{\Sigma \text{ Ratings in Rated Recommendations}}{\text{\# of Rated Recommendations}}$$

Cut off	Baseline	My Model
50	**1.43**	**1.42**
150	**1.32**	**1.51**
300	**1.29**	**1.70**
500	**1.25**	**1.78**

AVERAGE RATING BY CUTOFF
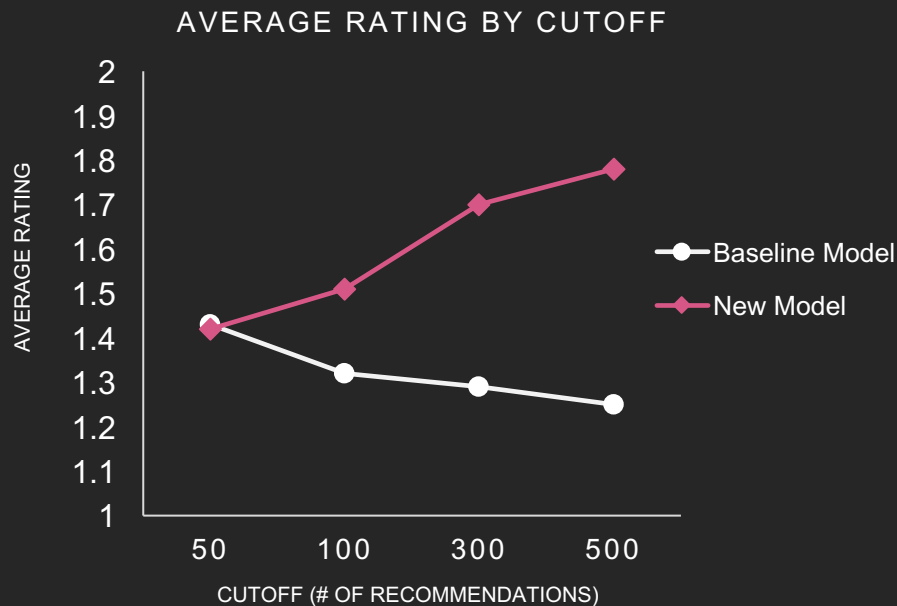
And yes …   A/B Test!

Future Work

- Natural language processing in profile and messages

- Include minority groups

- Image processing to recognize and reveal user's preferences

THANKS!

ANY QUESTIONS?

You can contact me at:

- tangjenny1224@gmail.com

- https://github.com/Jennytang1224