True Review

A Personalized Restaurant Recommender

Scope

PROBLEM

EXPLORE THE DATA

MODELING

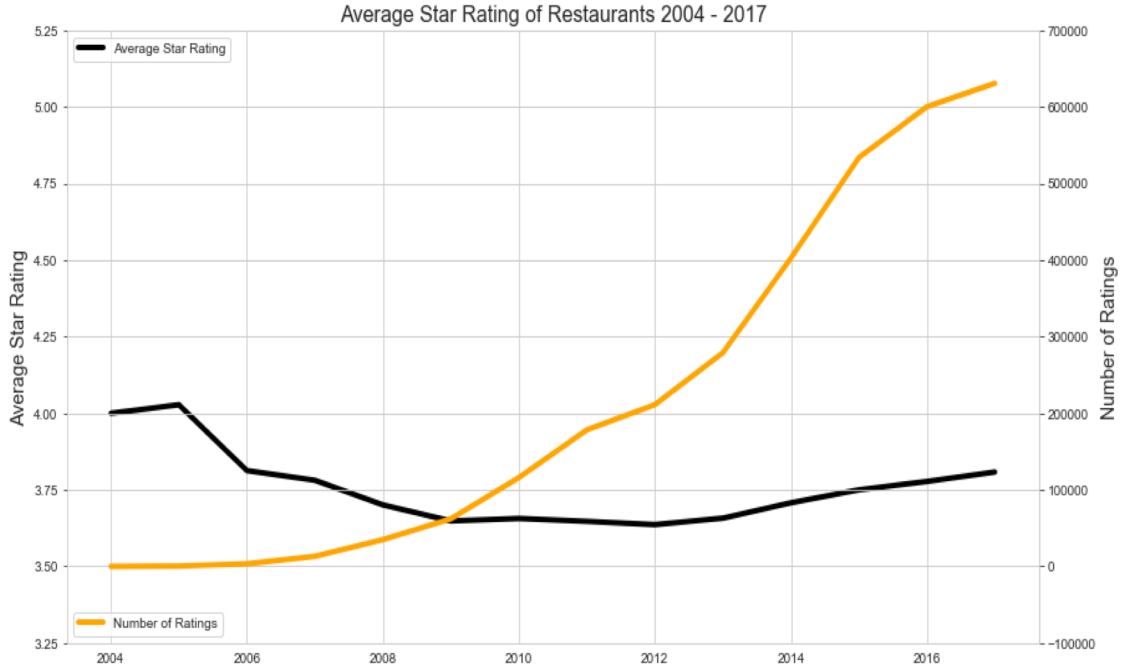
RESULTS

Problem

- What are users in Yelp reviews saying?
- Why should I trust Yelp reviewers, I'm different!?
- There are thousands of reviews for many restaurants, how can I read all of them to find similar users?
- How do I know which restaurants other users prefer?

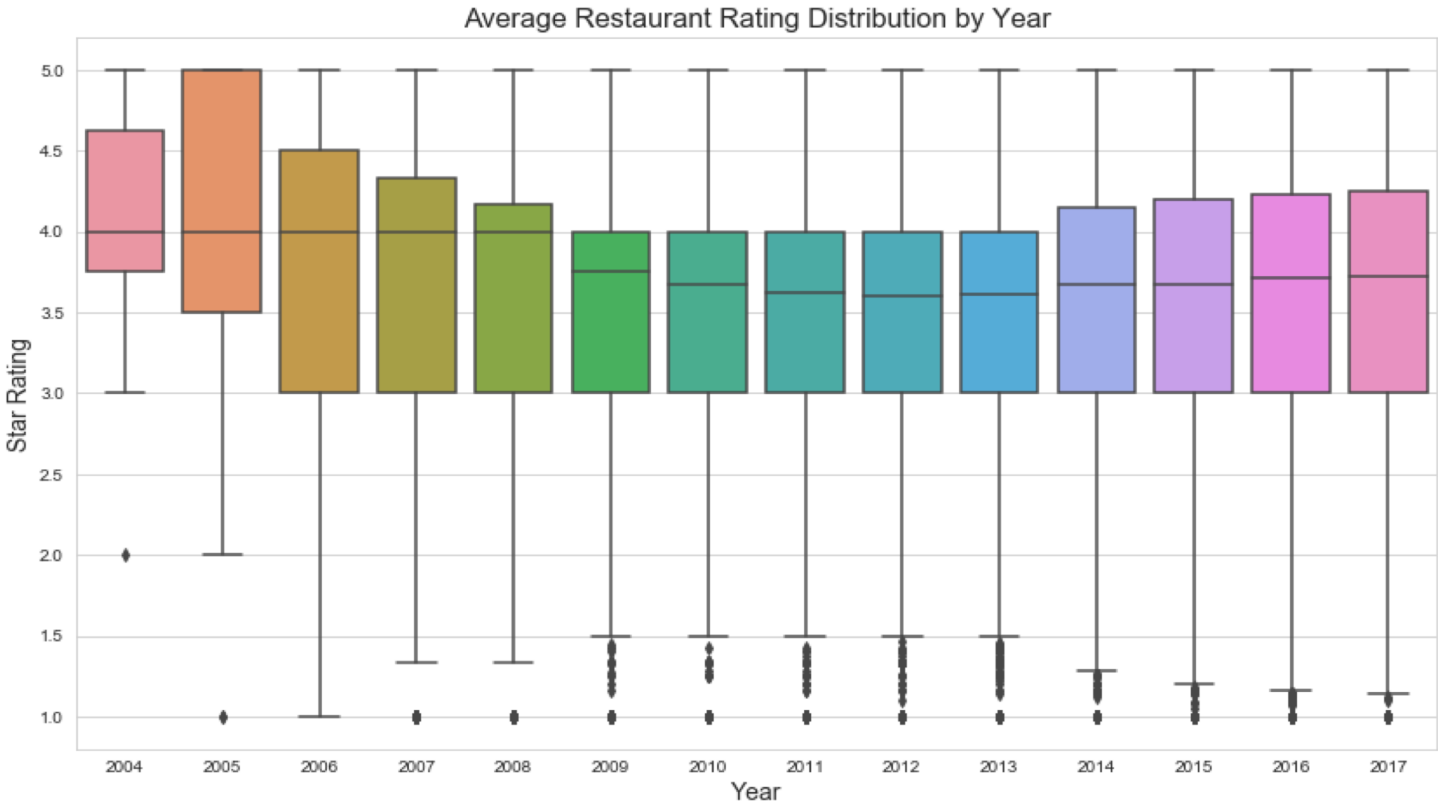
Explore the Data – Reviews Over Time

- Average star rating and number of reviews over the years



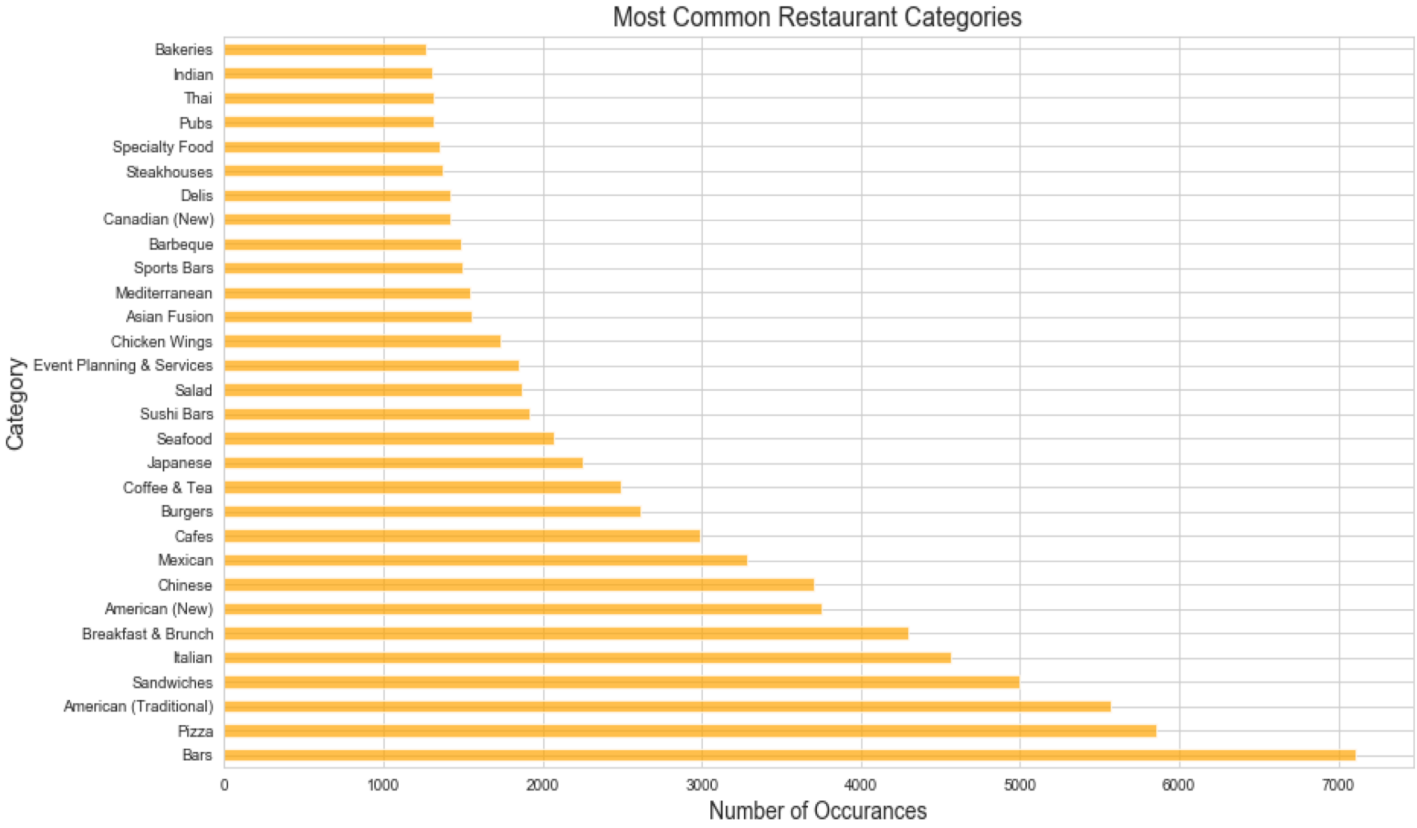
- As the number of reviews rises the average restaurant ratings appears to rise and start settling around 3.8 stars
- What does the distribution of reviews look like?

Explore the Data – Distribution of Ratings



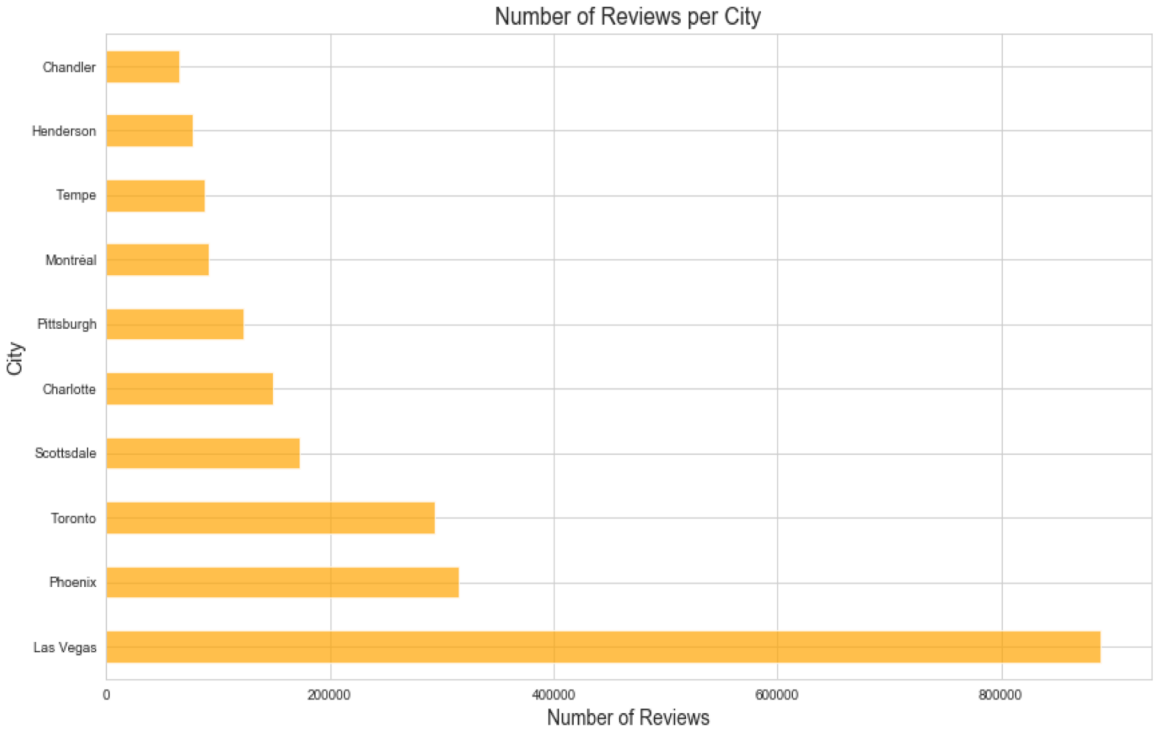
- Early ratings did not have many 1 stars, which increased the average and median.
- 1 star ratings still remain outliers. Note: this is a distribution of average restaurant ratings. Not many restaurants will have an average rating of 1 star.

Explore the Data – Common Categories



- Many Restaurant Categories are listed under in the Yelp database.
- These Categories were filtered and cleaned up. These are the most common and will be used to create content filtering system in the modeling stage.

Explore the Data – Number of Reviews



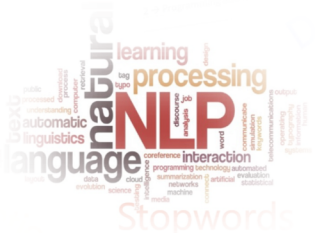
- Yelp Dataset had numerous cities.
- Picked Scottsdale for analysis due to relatively high review count and relatively dense user / restaurant matrix:

```
('Henderson', 0.145)
('Chandler', 0.14)
('Tempe', 0.12)
('Scottsdale', 0.076)
('Pittsburgh', 0.049)
('Charlotte', 0.048)
('Phoenix', 0.032)
('Montréal', 0.031)
('Las Vegas', 0.02)
('Toronto', 0.015)
```

Explore the Data – NLP

Random review for Eddie's House in Scottsdale, AZ:

'I really like this place. I have been numerous amt of times and I keep wanting more. The friendly bartenders, the exciting chef (Eddie). The best part of this place besides the good food and comfort level is their ALL Night Happy Hour. Yes All night $5 dollar specialty cocktails like an espresso martini or wines of the day. All first courses (apps) are half off too. The apps includes, lambchops ($19), Tuna tartar with wonton chips ($9) and so much more. Great s[ot and yearning for more since last night.'



- Remove symbols, characters, etc.
- Remove small words
- Form bigrams and trigrams

'numerous amt time keep want friendly bartender exciting chef good_part comfort level night dollar specialty_cocktail espresso_martini wine day first_course app half app include lambchop tartar wonton_chip much great yearning last_night'

Explore the Data – LDA

Random review for Eddie's House in Scottsdale, AZ:

'I really like this place. I have been numerous amt of times and I keep wanting more. The friendly bartenders, the exciting chef (Eddie). The best part of this place besides the good food and comfort level is their ALL Night Happy Hour. Yes All night $5 dollar specialty cocktails like an espresso martini or wines of the day. All first courses (apps) are half off too. The apps includes, lambchops ($19), Tuna tartar with wonton chips ($9) and so much more. Great s[ot and yearning for more since last night.'

- LDA Topics Extracted

dinner	0.21
happy hour, drinks	0.18
cheap, good, service	0.18
lunch	0.16
buffet	0.11
healthy	0.06

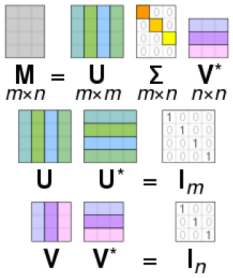
- Using LDA, we extract key topics from the review.
- Although not perfect, the higher probability topics are very accurate.
- These topics will be used to create a profile matrix

Modeling – Objective

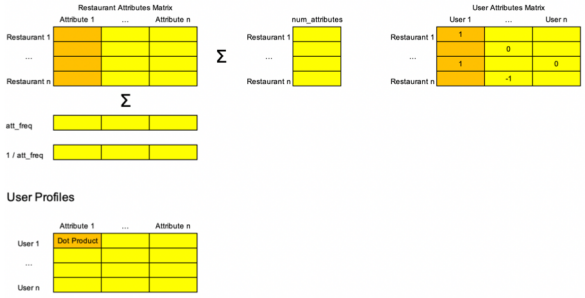
SVD

Content Filter

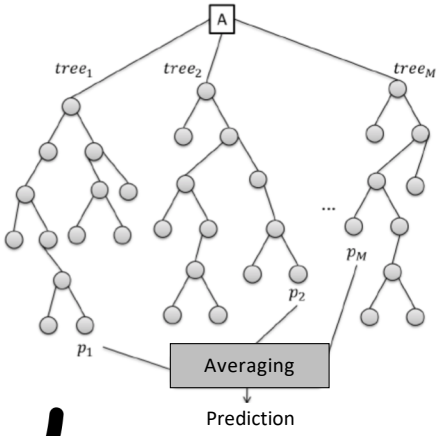
Stacked Ensemble



+



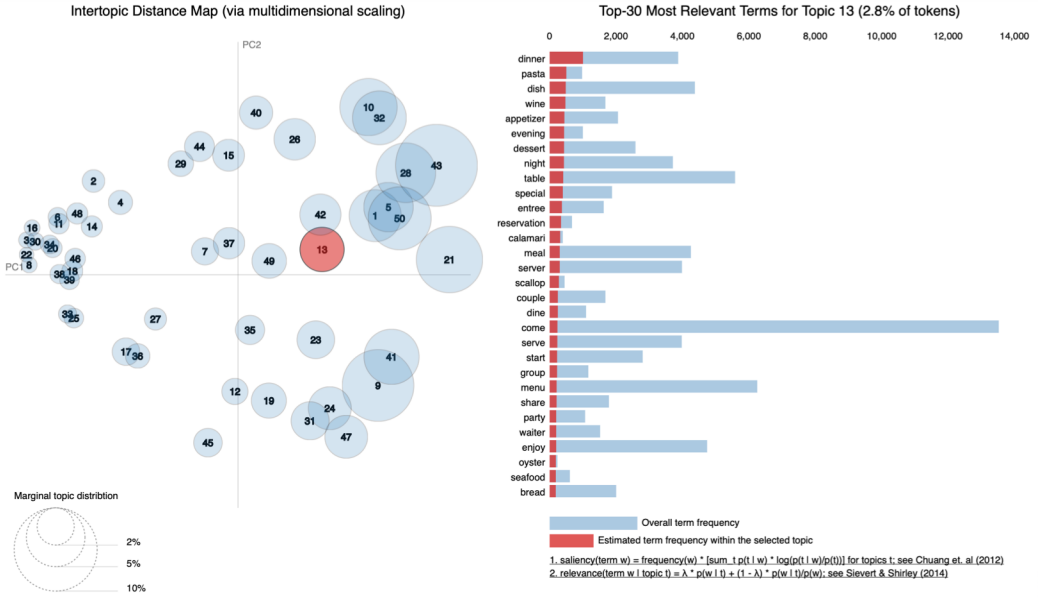
+



- Three approaches were used:
- SVD (Collaborative Filter)
- Content Filter
- Stacked Ensemble (ML Algorithms)

Final Rating

Modeling – LDA Topic Modeling



- LDA Topic Model was used to generate User Profiles utilized in the content filter
- Restaurant Categories (Yelp), Restaurant Attributes (Yelp), Restaurant Topics (LDA) were used to create a restaurant matrix.

Restaurant Attributes

Matrix 1 - Restaurant Categories

	Category 1	…	Category n
Restaurant 1			
…			
Restaurant n			

+

Matrix 2 - Retaurant Attributes

	Attribute 1	…	Attribute n
Restaurant 1			
…			
Restaurant n			

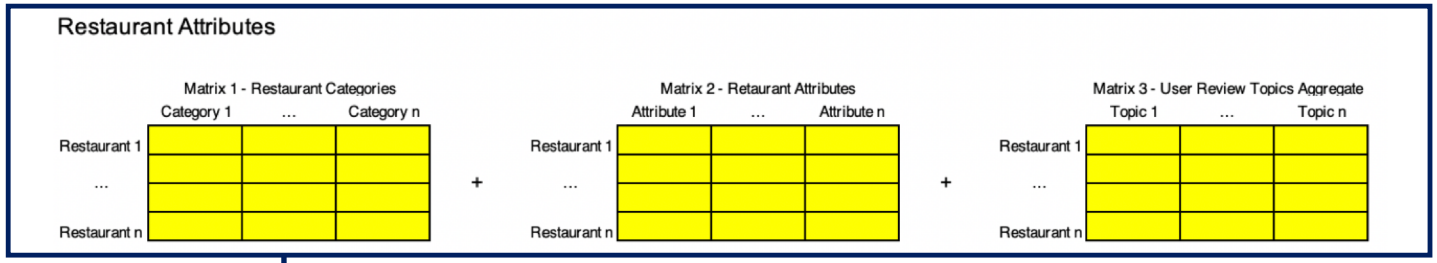
+

Matrix 3 - User Review Topics Aggregate

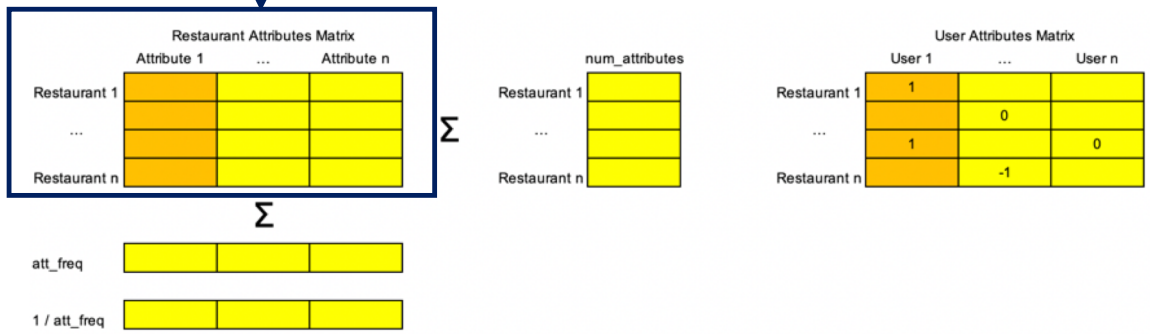
	Topic 1	…	Topic n
Restaurant 1			
…			
Restaurant n			

Modeling – Content Filter

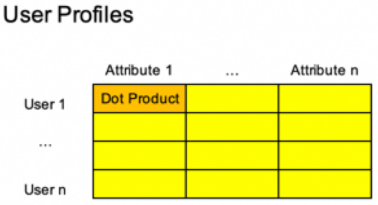
1



2

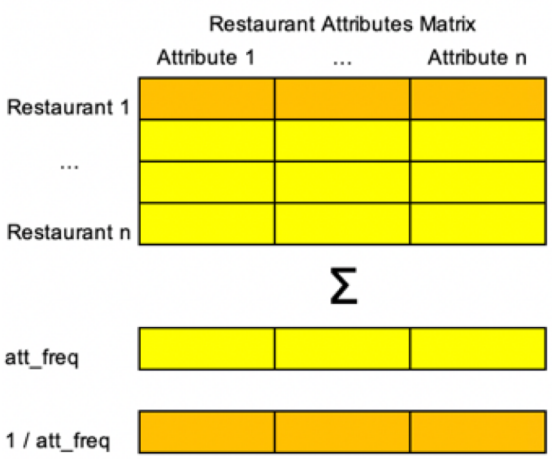


3

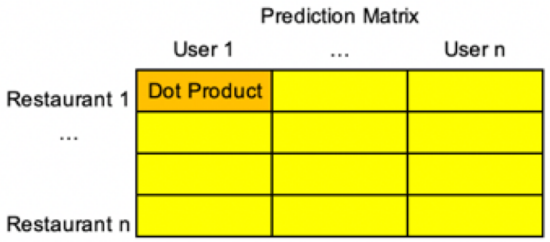


1. Create a normalized restaurant attributes matrix
2. Create a User attributes matrix (a User/Item matrix)
3. Form a User profile matrix by computing a similarity (dot product) between all restaurant attributes and all user attributes.

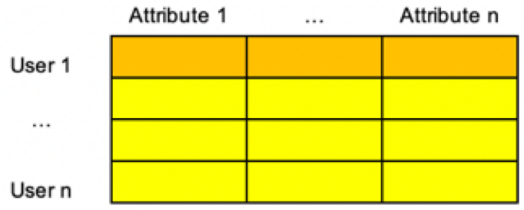
Modeling – Content Filter



1



User Profiles

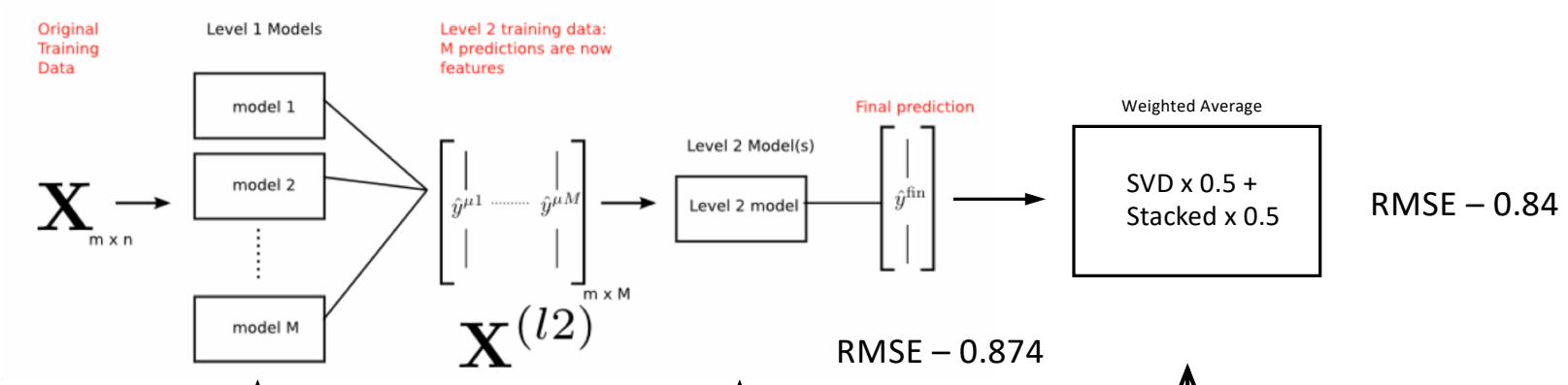


2

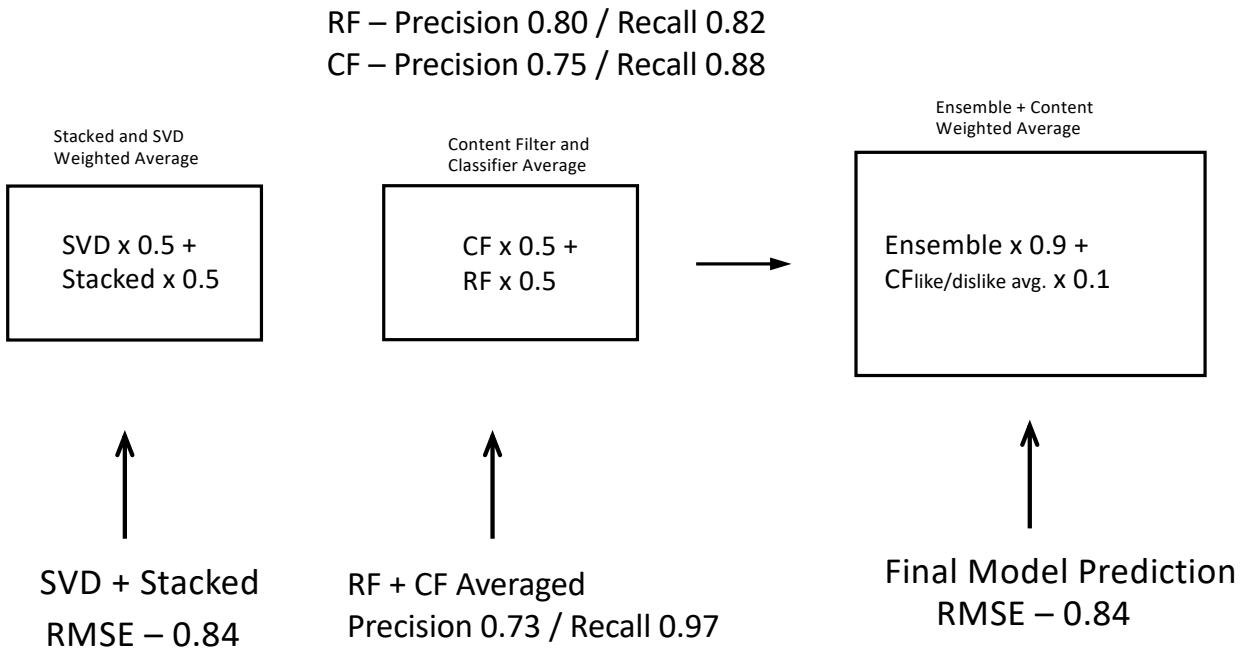
1. Compute the TF/IDF for Restaurant Attributes Matrix
2. Calculate the final prediction by computing the sum dot product of Restaurant attribute, TF/IDF, and User Attribute.

Final prediction will be a negative or positive number, indicating whether a user will dislike or like the restaurant, respectively.

Modeling – Stacked & Avg. Ensemble



Modeling – Combining Models



Results – Combined Predictor

RMSE Improvement	RMSE
Baseline – USER AVERAGE	0.90
Stacked Ensemble	0.874
SVD	0.878
SVD + Stacked Ensemble	0.84
SVD + Stacked Ens. + CF	0.84

Results – Next Steps

- Add Features – LDA Topic Model on entire dataset
- Tune model ensemble through GridSearch
- Extract latent features from SVD model
- Improve content filter algorithm