

Food★Findr

CONNECTING PEOPLE TO THE DISHES THEY LOVE

KEVIN WECHT

INSIGHT DATA SCIENCE FELLOW

2 FEBRUARY 2015

Mining Yelp reviews for sentiment of individual sentences

The Data:

Yelp Public Dataset (Phoenix, AZ)

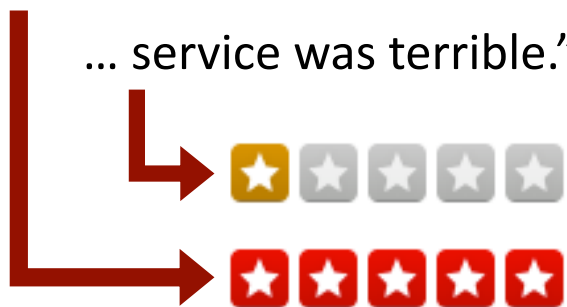
- 800 Mexican restaurants
- 40,000 reviews
- 300,000 sentences

The Idea:



"The guacamole was amazing...

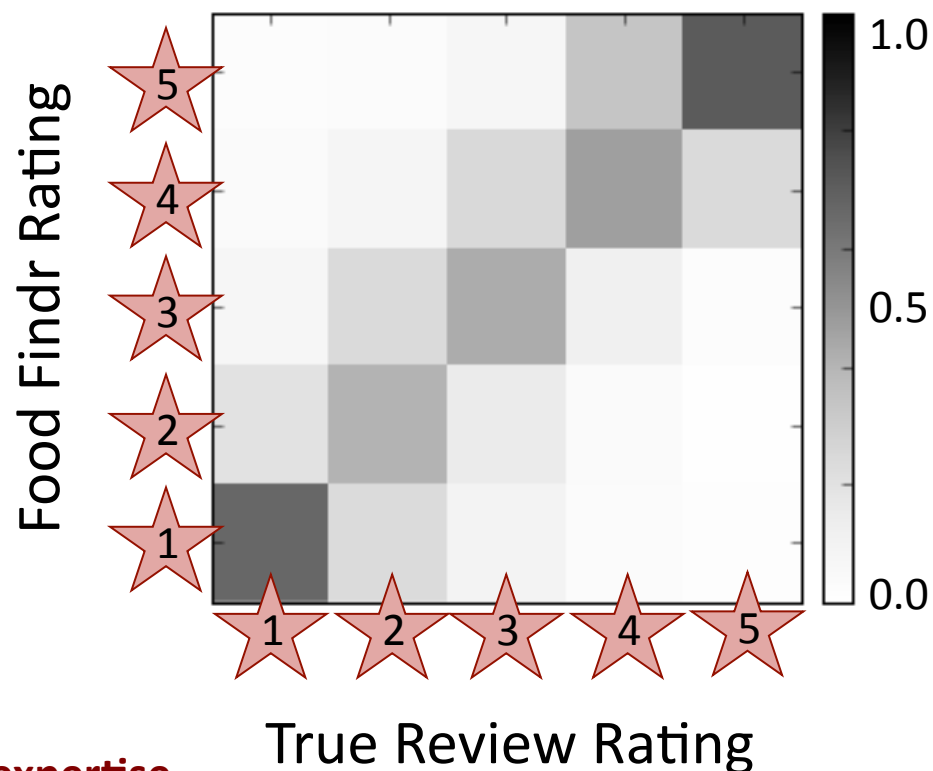
... service was terrible."



NLP: Lemmatization, negation handling, bi/tri-grams, TF-IDF

Classification: Multinomial Naïve Bayes

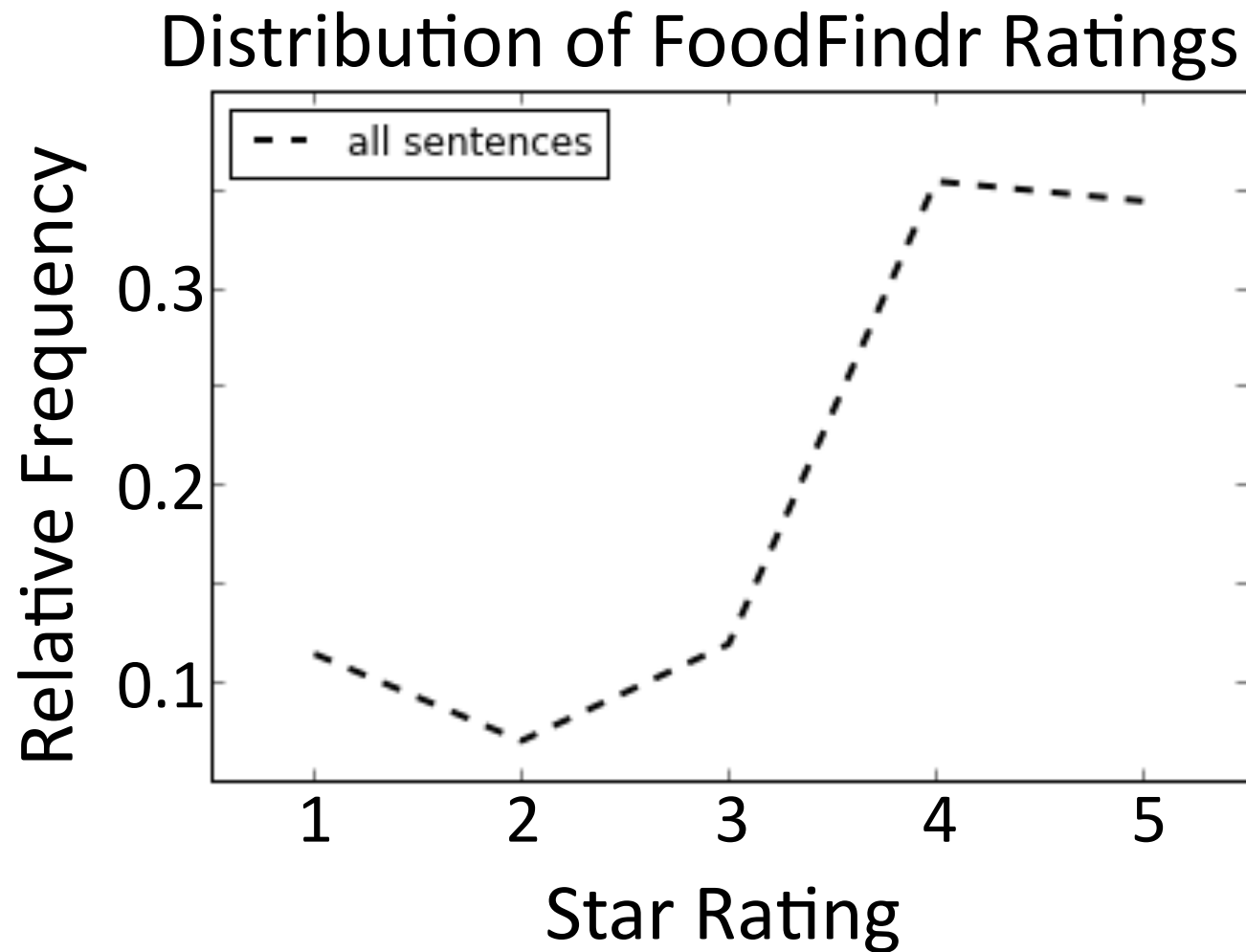
Classification of test data



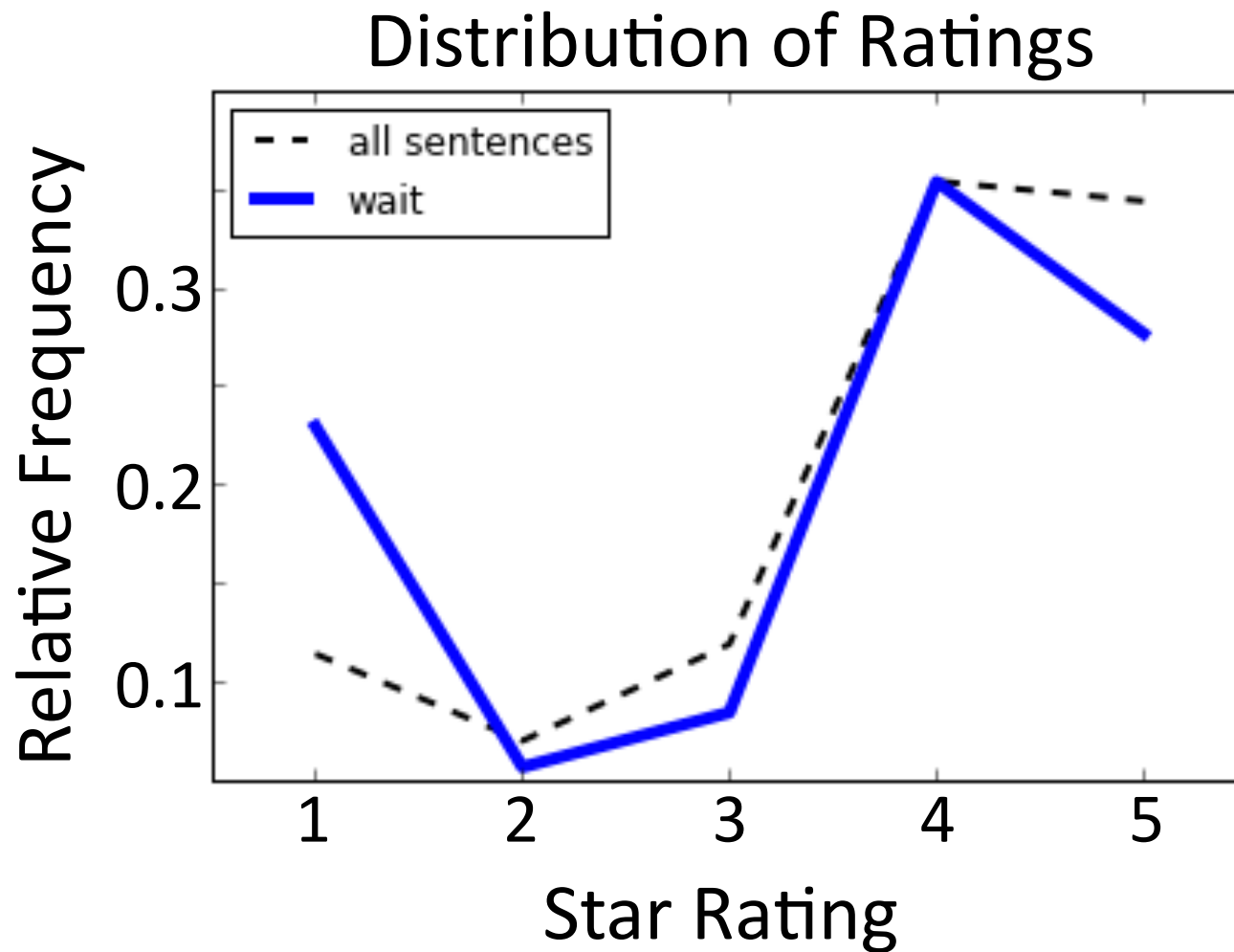
Scalable: feature creation requires no domain expertise.

Accuracy on 1-sentence reviews equals that of full test set --> Apply to individual sentences.

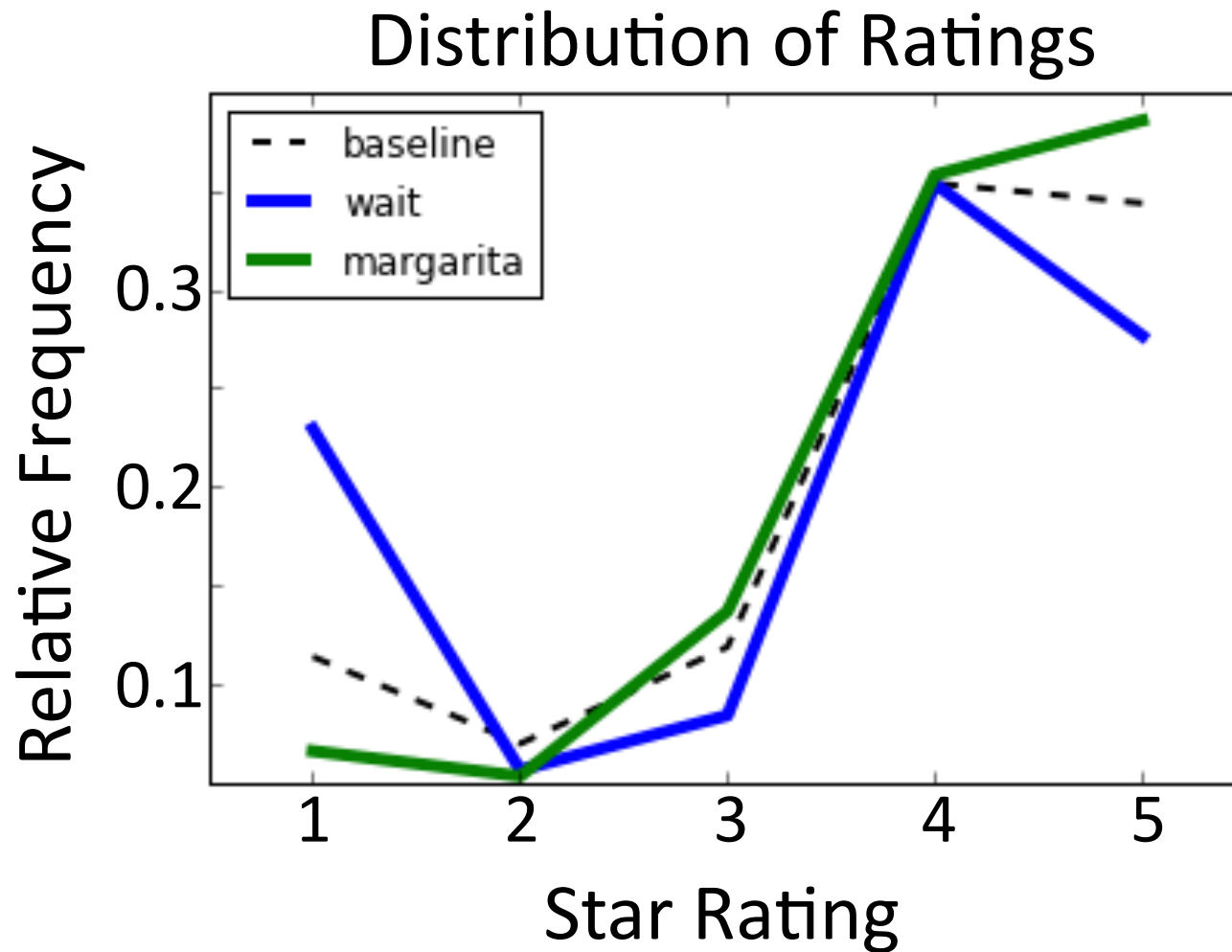
Extracting information from reviews



Extracting information from reviews



Extracting information from reviews



Food Findr extracts sentiments associated with common terms, providing a richer information experience to customers.

Examining a single business

This capability becomes particularly valuable to both customers and businesses when looking at a single restaurant.

Restaurant Name: Los Sombreros

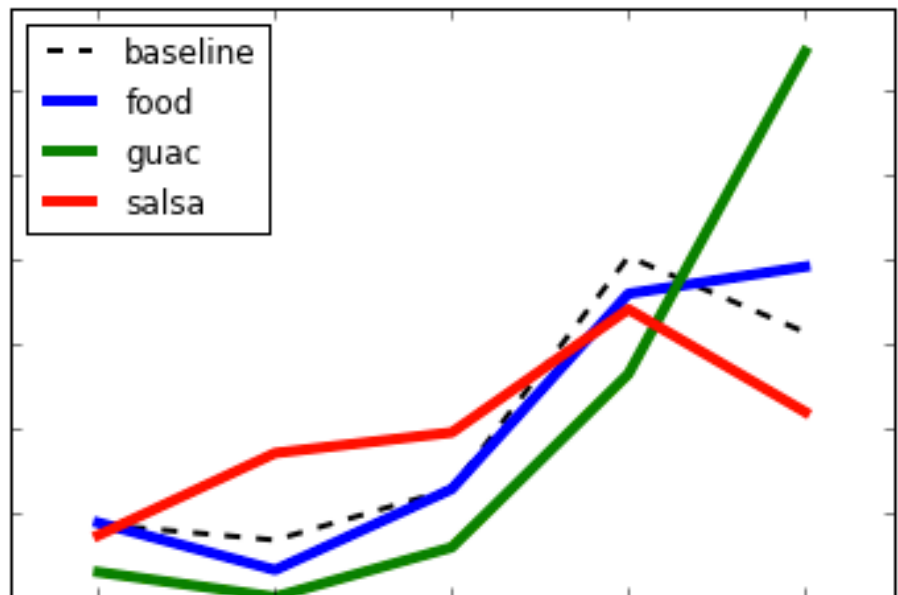
Category	Yelp Rating	FoodFindr Score
food	3.5	3.71604938272
atmosphere	3.5	4.21428571429
guac	3.5	4.40909090909
service	3.5	3.0
drinks	3.5	3.74418604651

Front page of Yelp reviews:

“The salsa was ok...

The guacamole was actually very good...

Tacos and chile relleno were mediocre.”



Extract and visualize actionable information regarding customer sentiment of a business's products.

Kevin Wecht

This page is about me