Word Salad
Relating Food Prices and Descriptions

Victor Chahuneau
Kevin Gimpel
Bryan R. Routledge
Lily Scherlis
Noah A. Smith

Outline

- Motivation
- Data
- Models

1. Item price prediction
2. Price range prediction
3. Joint price & sentiment prediction

14 juillet 2012
Chicken

2 pcs fried chicken meal

$4.99

Chicken Quesadillas
Made with fresh salsa, jack and cheddar cheese.

$6.99

Caesar Salad
Romain hearts, croutons, shaved parmesan cheese and classic Caesar dressing

$9.95

Poulet Cajun

$28.00
Noah's Ark Original Deli

Children's Menu

**MOUNT ARARAT**
Spaghetti & meatballs $8.95

**BIG BIRD**
Southern fried chicken served with french fries $9.95

**TENDER LOVE**
Breaded chicken tenders served with french fries $10.95

**TEENY WEENY**
½ Deli sandwich served with french fries $9.95

**GAN EDEN**
Grilled chicken breast served with french fries $10.95

Desserts

**HOMEMADE CHOCOLATE CHIP COOKIES** $3.95

**JELLO** $3.95

**BROWNIE** $4.95

Noah's Ark Original Deli

Categories: Delis, Kosher

399 Grand St
(between Suffolk St & Clinton St)
New York, NY 10002
Neighborhood: Lower East Side

Price Range: $$
Good for Kids: Yes
Takes Reservations: Yes
Delivery: Yes
Take-out: Yes
Waiter Service: Yes
Outdoor Seating: No

⭐⭐⭐⭐

Love Noah's Ark! Great comfort food. The ambiance is very deli style. Everything I've had here has been very good - from the burgers to the matzo ball soup. As a southerner, I must say it's quiet exciting for me to be able to get kosher chicken fried steak. It's delicious! Their burger menu is also quiet extraordinary!

⭐⭐⭐⭐⭐

Ever have a super indecisive moment when faced with a somewhat extensive online menu and needing to order food?? I feel like it's so much pressure and I usually order the wrong thing half the time. Sometimes it's also what you're craving or feel like you're craving at that moment while faced with the daunting task...

⭐⭐⭐⭐⭐

This place is kosherific. Matzyahu is their custy and you will see a pic of him eating there on the door. I don't keep kosher so I wouldn't come here for a burger or anything else that would be better with cheese. I would come here for the pastrami, corned beef, brisket, and knishes.
Data - facts

- 7 American cities
  - New York
  - San Francisco
  - Chicago
  - Los Angeles
  - Boston
  - Philadelphia
  - Washington DC

- 12k restaurants
Data - facts

- 900k menu items
- 400k distinct names

- 1.5M reviews
- 80M words
Models

- Linear
  - linear regression
  - logistic regression

- Features: binary / bag of words

- Sparsity: $l_1$ regularization
1 - Item price prediction

**MenuNames**
- soggy fries
- n/fries+soggy
- soggy fries
- soggy fries

**MenuDesc**
- drenched in brown gravy
- drenched
- brown
- brown gravy
- drenched in brown

**Metadata**
- Category: deli
- Category: kosher
- TakesReservations: yes
- Attire: casual
- GoodForKids: yes
- HasTV: no

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th>MRE</th>
<th>number features</th>
<th>model size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predict mean</td>
<td>3.70</td>
<td>43.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Predict median</td>
<td>3.67</td>
<td>43.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAME baseline</td>
<td>3.66</td>
<td>45.64</td>
<td>267,945</td>
<td>240,139</td>
</tr>
<tr>
<td><strong>Metadata</strong></td>
<td>3.55</td>
<td>43.11</td>
<td>268,450</td>
<td>258,828</td>
</tr>
<tr>
<td><strong>MenuNames</strong></td>
<td>3.23</td>
<td>38.33</td>
<td>896,631</td>
<td>230,840</td>
</tr>
<tr>
<td>+ <strong>MenuDesc</strong></td>
<td>3.19</td>
<td>36.23</td>
<td>1,981,787</td>
<td>151,785</td>
</tr>
<tr>
<td>+ <strong>Metadata</strong></td>
<td>3.08</td>
<td>34.97</td>
<td>1,982,363</td>
<td>148,774</td>
</tr>
</tbody>
</table>
Authenticity in America
Class Distinctions in Potato Chip Advertising

J. Freedman and D. Jurafsky. 2011
### MenuDesc - Authenticity

<table>
<thead>
<tr>
<th>Term</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>old time favorite</td>
<td>-0.112</td>
</tr>
<tr>
<td>fashioned</td>
<td>-0.034</td>
</tr>
<tr>
<td>line caught</td>
<td>-0.028</td>
</tr>
<tr>
<td>all natural</td>
<td>-0.028</td>
</tr>
<tr>
<td>traditional</td>
<td>-0.009</td>
</tr>
<tr>
<td>natural</td>
<td>3e-4</td>
</tr>
<tr>
<td>classic</td>
<td>0.002</td>
</tr>
<tr>
<td>free range</td>
<td>0.004</td>
</tr>
<tr>
<td>real</td>
<td>0.004</td>
</tr>
<tr>
<td>fresh</td>
<td>0.006</td>
</tr>
<tr>
<td>homemade</td>
<td>0.010</td>
</tr>
<tr>
<td>authentic</td>
<td>0.012</td>
</tr>
<tr>
<td>organic</td>
<td>0.020</td>
</tr>
<tr>
<td>specialty</td>
<td>0.025</td>
</tr>
<tr>
<td>special</td>
<td>0.033</td>
</tr>
<tr>
<td>locally</td>
<td>0.037</td>
</tr>
<tr>
<td>natural grass fed</td>
<td>0.038</td>
</tr>
<tr>
<td>artisanal</td>
<td>0.064</td>
</tr>
<tr>
<td>raised</td>
<td>0.066</td>
</tr>
<tr>
<td>heirloom</td>
<td>0.083</td>
</tr>
<tr>
<td>wild</td>
<td>0.084</td>
</tr>
<tr>
<td>hormone</td>
<td>0.085</td>
</tr>
<tr>
<td>farmed</td>
<td>0.099</td>
</tr>
<tr>
<td>hand picked</td>
<td>0.101</td>
</tr>
<tr>
<td>wild caught</td>
<td>0.116</td>
</tr>
<tr>
<td>farmhouse</td>
<td>0.133</td>
</tr>
</tbody>
</table>
**MenuDesc - size**

<table>
<thead>
<tr>
<th>Value</th>
<th>Description</th>
<th>of chicken</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.102</td>
<td>slices _</td>
<td></td>
</tr>
<tr>
<td>-0.032</td>
<td>bits _</td>
<td></td>
</tr>
<tr>
<td>-0.030</td>
<td>cubes _</td>
<td></td>
</tr>
<tr>
<td>-0.024</td>
<td>pieces _</td>
<td></td>
</tr>
<tr>
<td>-0.001</td>
<td>strips _</td>
<td></td>
</tr>
<tr>
<td>0.015</td>
<td>chunks _</td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>morsels _</td>
<td></td>
</tr>
<tr>
<td>0.040</td>
<td>pcs _</td>
<td></td>
</tr>
<tr>
<td>0.042</td>
<td>cuts _</td>
<td></td>
</tr>
</tbody>
</table>
### MenuDesc - style

<table>
<thead>
<tr>
<th>Method</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>panfried</td>
<td>-0.094</td>
</tr>
<tr>
<td>chargrilled</td>
<td>-0.029</td>
</tr>
<tr>
<td>cooked</td>
<td>-0.012</td>
</tr>
<tr>
<td>boiled</td>
<td>-0.006</td>
</tr>
<tr>
<td>fried</td>
<td>-0.005</td>
</tr>
<tr>
<td>steamed</td>
<td>0.011</td>
</tr>
<tr>
<td>charbroiled</td>
<td>0.015</td>
</tr>
<tr>
<td>grilled</td>
<td>0.022</td>
</tr>
<tr>
<td>simmered</td>
<td>0.025</td>
</tr>
<tr>
<td>roasted</td>
<td>0.034</td>
</tr>
<tr>
<td>sauteed</td>
<td>0.034</td>
</tr>
<tr>
<td>broiled</td>
<td>0.053</td>
</tr>
<tr>
<td>seared</td>
<td>0.066</td>
</tr>
<tr>
<td>braised</td>
<td>0.068</td>
</tr>
<tr>
<td>stirfried</td>
<td>0.071</td>
</tr>
<tr>
<td>flamebroiled</td>
<td>0.106</td>
</tr>
</tbody>
</table>
## 2- Price range prediction

**Reviews**
- decided
- get
- cajun fries
- cajun fries decided to get
- get the cajun

- no deli cheap
- no deli
- deli is cheap

**Metadata**
- Category: deli
- Category: kosher
- Neighborhood: Lower East Side
- TakesReservations: yes
- Attire: casual
- GoodForKids: yes
- HasTV: no

<table>
<thead>
<tr>
<th></th>
<th>MAE</th>
<th>Acc.</th>
<th>number features</th>
<th>model size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predict mode</strong></td>
<td>0.54</td>
<td>48.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Menu</strong></td>
<td>0.39</td>
<td>66.29</td>
<td>1,910,622</td>
<td>995</td>
</tr>
<tr>
<td><strong>Metadata</strong></td>
<td>0.24</td>
<td>77.15</td>
<td>591</td>
<td>219</td>
</tr>
<tr>
<td><strong>Reviews</strong></td>
<td>0.22</td>
<td>79.76</td>
<td>3,027,470</td>
<td>1,567</td>
</tr>
<tr>
<td><strong>+ Metadata</strong></td>
<td>0.21</td>
<td>80.36</td>
<td>3,027,943</td>
<td>1,376</td>
</tr>
</tbody>
</table>

**Price range:**
- $$$$$
- $$$
- $$
- $
2- Polarity prediction

Accuracy: 87%

B. Pang and L. Lee 2008
Opinion mining and sentiment analysis
2/good service + 0.55
1/service +15.13

great dark, sexy atmosphere. good service. nice variety of tasty cocktails. sushi tasted good,

this place gets 5 stars for food, drinks, and atmosphere

... and negative stars for the jar-headed douche bag `` security guards ''

2/negative stars -19.77
1/stars - 3.00

$ price
★ sentiment
3- Joint $★$ prediction

**REVIEWS**
- decided
c- get
c- cajun
c- fries
c- cajun fries
decided to get
g- get the cajun
- no
c- deli
c- cheap
- no deli
c- deli is cheap

$\begin{array}{|c|c|}
\hline
\$ & \$\\
$ & $\\
\hline
\end{array}$
terrible
awful
disgusting
overpriced
worse
not worth
mediocre
pasta
horrible
rude
amazing
love this place
rice
can't wait
incredible
edible
great food
calamari
awesome
tasteless
highly recommend
can't wait
highly
n't wait
hands down
never go back
n't beat
great
weed
favorite
place is great
fish scallops
low price
taste
less than
wine
excellent
tasted like
taste
fish
service
wait staff
taste like
wine list
wine
food
lukewarm
taste
waste your
tasty
bad reviews
favorite
new favorite
money
hot
beef
after
sandwiches
ketchup
food
clam chowder
taste
walked in
ravioli
taste
n't beat
taste
sick
food
mussels
taste
fried rice
food
fish
taste
-service
waste
food
food
food
food
three stars
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
food
fo
We conjecture that systematic trends in errors can help detect patterns in restaurant prices. For example, if errors correlate with geography, we may identify regions with higher- or lower-than-expected prices. To test this for New York, we retrained our price range predictor using the entirety of New York as the test set, Philadelphia as the dev set, and the remaining cities as training data. We only used features from the reviews, omitting geographic metadata features. We then compare the predictions of the model to the real price range values in geographic context. More precisely, for every restaurant represented by its projected coordinates, we compute the relative error $\hat{y} - y$ at this point and smooth the resulting two-dimensional price distribution using a triangular kernel. We obtain the map of Manhattan shown in Figure 1, revealing trends in our predictions.

We first note the prevalence of blue in the plot; since we trained our models on cheaper cities and tested on Manhattan, we are systematically underpricing Manhattan restaurants. We see the most severe underprediction in Midtown East, an area known for its expensive cost of living. However, some areas differ from the overall trend; we observe moderate overprediction in Chinatown and parts of Greenwich Village and the East Village, areas which are generally known to be lower-priced than most of lower Manhattan.

Figure 1: Smoothed distribution of relative differences between the value predicted from the reviews and the real price range for 2965 restaurants in Manhattan.

We clarify that the prediction is not made directly using explicit mentions by users in their reviews (e.g., uses of key phrases such as "overpriced" and "great deal") but rather by considering predictions from a regression model using features aggregated across all reviews for the restaurant. While a simpler technique based on counts of key phrases may also be able to locate geographic trends such as these, our approach leverages more sophisticated modeling techniques.
Bon appétit !