Where Not to Eat? Improving Public Policy by Predicting Hygiene Inspections Using Online Reviews

Stony Brook University

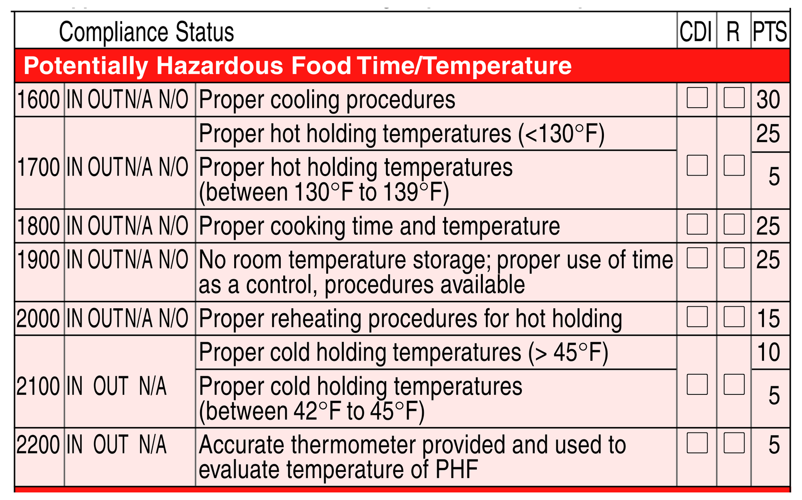
Jun Seok Kang*, Polina Kuznetsova*, Michael Luca**, Yejin Choi*

* {junkang, pkuznetsova, ychoi} @cs.stonybrook.edu

** mluca@hbs.edu

Harvard Business School

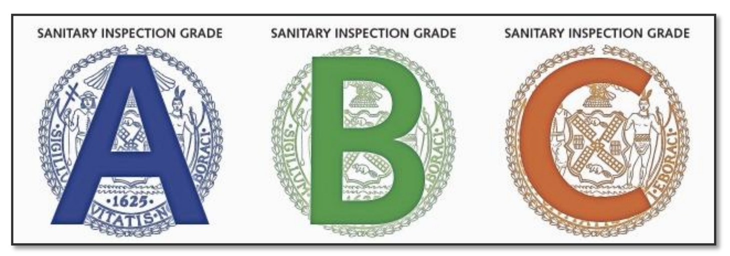
Motivation



Compliance Status			CDI	R	PTS
Potentially Hazardous Food Time/Temperature					
1600	IN OUT N/A N/O	Proper cooling procedures			30
		Proper hot holding temperatures (<130°F)			25
1700	IN OUT N/A N/O	Proper hot holding temperatures (between 130°F to 139°F)			5
1800	IN OUT N/A N/O	Proper cooking time and temperature			25
1900	IN OUT N/A N/O	No room temperature storage; proper use of time as a control, procedures available			25
2000	IN OUT N/A N/O	Proper reheating procedures for hot holding			15
		Proper cold holding temperatures (> 45°F)			10
2100	IN OUT N/A	Proper cold holding temperatures (between 42°F to 45°F)			5
2200	IN OUT N/A	Accurate thermometer provided and used to evaluate temperature of PHF			5

King County(Seattle)'s inspection form(partial)

- Many counties & cities such as NYC or LA require restaurants **to post their inspection grades.**



- This grade posting affects the revenue of the restaurant substantially, and also reduces the risk of food-borne illness (e.g., Jin and Leslie (2003), Henson et al. (2006))
- But these *health departments* have *limited resources to dispatch inspectors.*
  - Among the Seattle restaurants listed on Yelp.com (2006~2013), **more than 50% *of them didn't have an inspection record!***
- Instead, we have hundreds of thousands of reviews of those places written by the very people who actually have visited them!

➔ Prediction of the hygiene status of restaurants by using online reviews
Then, *is this a sentiment analysis task?* **No.**

Related Work

- Public health surveillance by tracking textual signals in micro-blogs
  - *Aramaki et al. (2011), Sadilek et al. (2012b), Sadilek et al. (2012a), Sadilek et al. (2013), Lamb et al. (2013), Dredze et al. (2013), von Etter et al. (2010)*

  ➔ Our work differs in two ways

  1) Examines all words to predict the inspection outcome by correlating seemingly distant terms & concepts, instead of focusing on specific keywords for specific illness.

  2) First attempt to examine online reviews to improve public policy
- Our work draws from prior studies on online reviews for sentiment analysis *(e.g., Pang and Lee (2008))* and deception detection (e.g., *Mihalcea and Strapparava (2009), Ott et al. (2011), Feng et al. (2012))*
- Also, previous studies for aspect-based sentiment analysis (e.g., *Titov and McDonald (2008), Brody and Elhadad (2010), Wang et al. (2010)*) would be a fruitful venue for further investigation.

Data

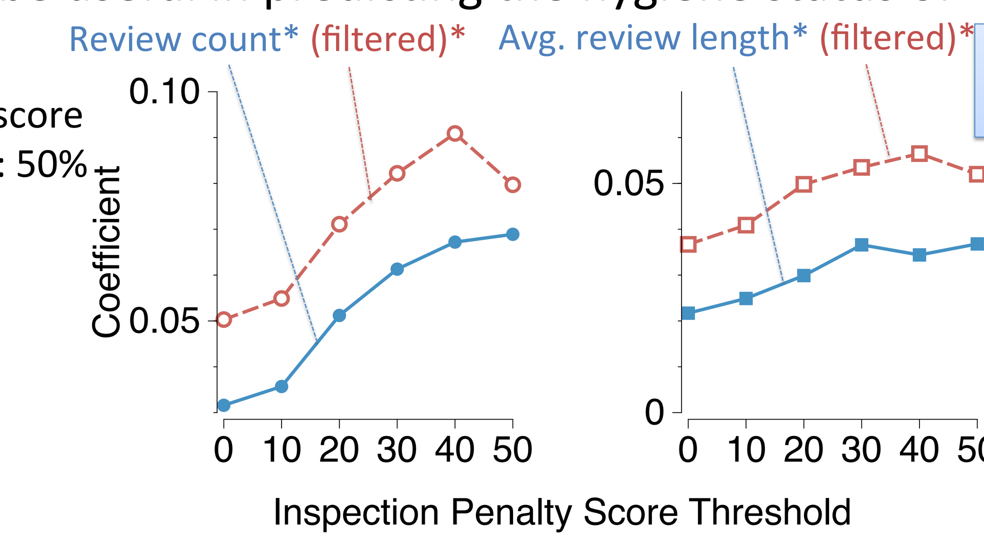
- Yelp.com Reviews (Seattle): 2006 ~ 2013 January
  - Available at http://cs.stonybrook.edu/~junkang/hygiene
- Inspection Records (Seattle, King County): 2006~ 2013
  - Available at http://data.kingcounty.gov

➔ 13k inspections: 1756 restaurants with 152k reviews

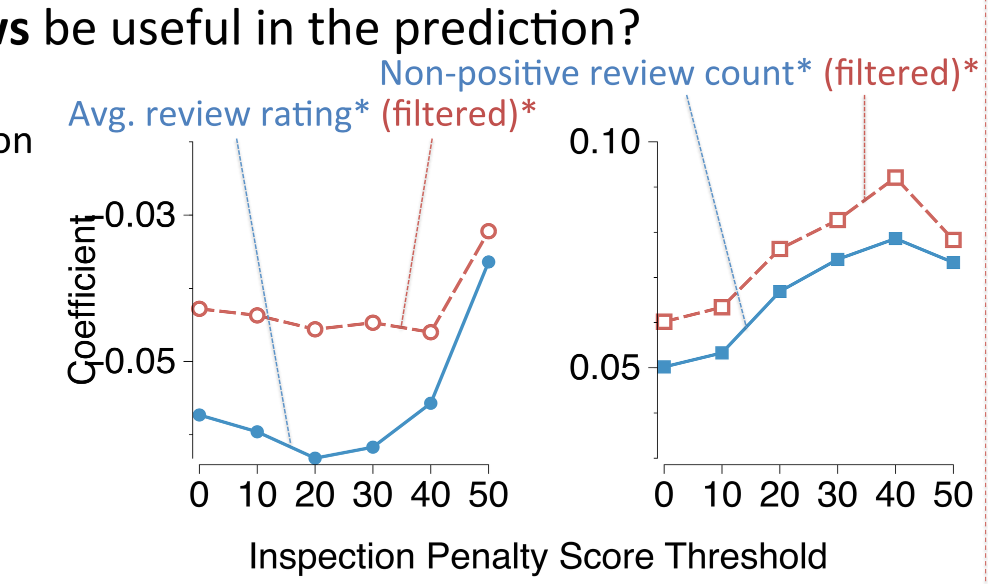
Correlates of Inspection Score

(Spearman Coefficients. Labeled '*' for statistical significance (p ≤ 0.05))

- Would the **Volume of Reviews** be useful in predicting the hygiene status of the restaurants?
  - Mild positive correlation to inspection score
  - Classification result using review count: 50%

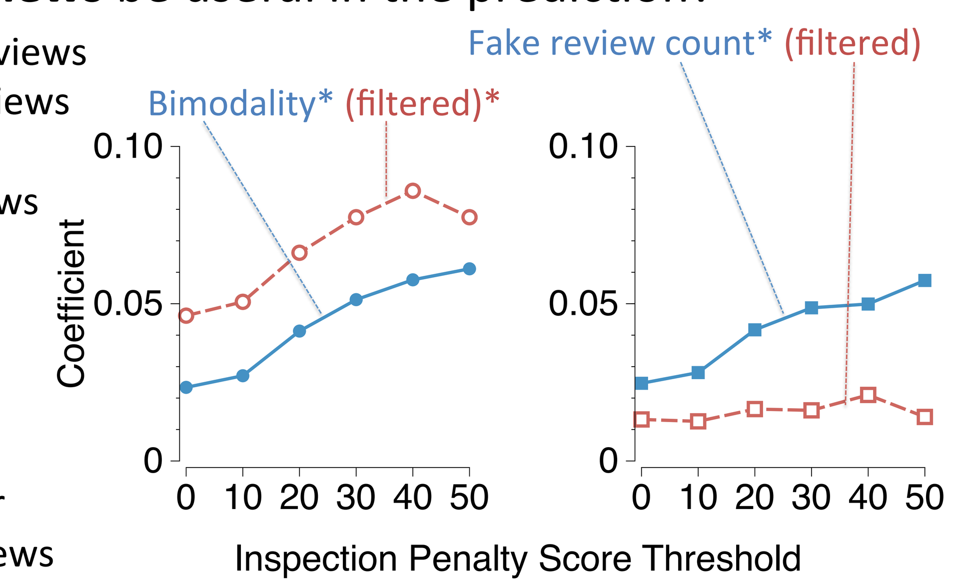


- Would the **Sentiment of Reviews** be useful in the prediction?
  - Mild correlations to inspection score
    - *Avg. review rating*: negative correlation
    - *Non-positive review count*: positive correlation
  - Classification Result:
    - **Avg. review rating: 57.52%**
    - Non-positive review count: 52.94%



* Filtered reviews that are far from the average rating (> 2.0)
- To observe the impact of dubious, potentially deceptive reviews
- Filtered reviews show ***stronger correlation*** suggesting *the existence of deceptive reviews covering unhappy customers.*

- Would the **Deceptiveness of Reviews** be useful in the prediction?
  - *Unhygienic restaurants* with negative reviews have good motivation to solicit fake reviews
  - Also, *assiduous but hygienic restaurants* may solicit fake reviews for better reviews
  - ***Bimodality***
    - Distribution of ratings as the sign of deceptive review activities (Feng et al. (2012))
  - ***Fake review count***
    - Classified using the deception classifier trained on a set of fake & truthful reviews following Ott et al. (2011)
  - Mild positive correlation to inspection score

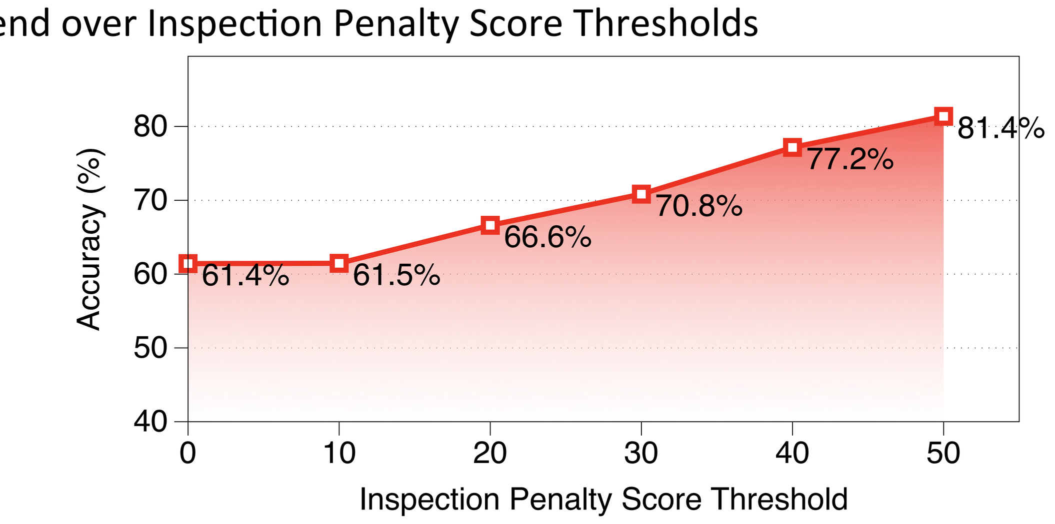


Content-based Predictions

- Would the **Content of Reviews** be useful in the prediction?
  - Reviewers are *not instructed to write about the hygiene status of the restaurants*, and they *often do not even mention* about it but other facts

    ➔ ***Noisy data***
  - Most of the customers *do not know the inspection items in detail*, and usually *do not have an access to the key areas.*

    ➔ ***Insufficient information***
  - Previous works in health surveillance monitor influenza or food-poisoning outbreaks from micro blogs.
- *Features based on customer's opinion*
  - Aggregated opinions: **average review rating**
  - Content of reviews: **unigram, bigram**
- *Features based on meta data*
  - Cuisine: Yelp's restaurant **category**
  - Location: **ZIP code**
  - Inspection History: **avg. penalty score, previous inspection result**
  - **Review count & Non-positive review count**
- *Classification Results using SVM & SVR (liblinear, Fan et al., 2008)*
  - Trend over Inspection Penalty Score Thresholds



- Feature Compositions

Type	Features	SVM	SVR	
		Accuracy	MSE	SCC
Meta Data	Cuisine	*66.2	0.23	0.15
	zip code	*67.3	0.21	0.17
	Inspection history	*72.2	0.20	0.20
Texual Contents	Unigram	78.4	0.46	0.10
	Bigram	*76.6	0.48	0.05
	Unigram + Bigram	82.7	0.44	0.10
All Features		81.4	0.19	0.26

(MSE: Mean Squared Errors, SCC: Squared Correlation Coefficients)

Insightful Cues

Sentiment words represent how reviewers felt
- Unhygienic: *cheap, never*
- Hygienic: *lovely, yummy, generous, friendly, great, nice*

Hygiene related words are overwhelmingly negative. Reviewers do complain when the restaurants are dirty (*Gross, mess, sticky, smell, restroom, dirty)*, but do not seem to complement on the cleanliness. Instead, they focus on other positive aspects, such as **details of food**, **atmosphere**, and their **social occasions**.

Whom & When represent the social occasion which reviewers seems to focus on if the restaurants are clean enough.
- Hygienic: *date, weekend, our, husband, evening, night*

Details of Food differ between the hygienic & unhygienic reviews. For unhygienic places, people talk about the ingredients of the food. If the place is clean, they talk about how food is prepared or decorated.
- Unhygienic: *beef, pork, noodle, egg, soy, ramen, pho,*
- Hygienic: *brew, frosting, grill, crush, crust, taco, burrito, toast, celery, calamity, wine, broccoli, salad, flatbread, olive, pesto*

Service and Atmosphere correlate to inspection outcome even though they are not immediately related to hygiene status.
- Unhygienic: *door, student, sticker, or the size*
- Hygienic: *selection, attitude, atmosphere, ambiance, pretentious*

Cuisines are clearly correlated to the inspection outcome.
- Unhygienic: *Vietnamese, Dim Sum, Thai, Mexican, Japanese, Chinese, American, Pizza, Sushi, Indian, Italian, Asian*
- Hygienic: *Breakfast, Fish & Chips, Fast Food, German, Diner, Belgian, European, Sandwiches, Vegetarian*