The image contains a page from a document discussing the relationship between online reviews and hygiene inspections in restaurants. The text is divided into several sections, each with a title and paragraphs of text. The page includes tables, figures, and graphs that illustrate the correlation between certain features and inspection scores. The document covers topics such as the use of sentiment analysis to predict hygiene inspection outcomes, the importance of public health in restaurant inspections, and the role of online reviews in evaluating restaurant quality.

**Motivation**
- Many counties and cities across the U.S. require restaurants to post their inspection grades.

**Related Work**
- Public health surveillance by tracking textual data in micro-blogs.

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

**Insightful Cues**

**Content-based Predictions**
- Would the Volume of Reviews be useful in predicting the hygiene status of the restaurants?
- Would the Sentiment of Reviews be useful in the prediction?
- Would the Deceptiveness of Reviews be useful in the prediction?

**Correlates of Inspection Score**
- Mild positive correlation to inspection score
- Non-positive review count: positive correlation
- Non-positive review count: 52.94%

**Inspections**
- Certain inspection results show stronger correlations suggesting the existence of deceptive reviews covering unhappy customers.

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

**Key takeaways**
- Reviews are not instructed to write about hygiene status of restaurants.
- Most of the customers do not know inspection items in detail, and usually do not have access to the key areas.

**Inspection Penalty Score Threshold**
- Consideration results using SVM & SVR (liblinear, Fan et al., 2008)
- Trend over Inspection Penalty Score Thresholds

**Insightful Cues**

- Unlabeled comments should represent social centers which reviewers seem to focus on if the restaurants are clean enough.
- Hygiene: date, weekend, our, husband, evening, night

**Hygiene**
- Review content
- Relevant words on unhygienic negative
- Details of Food differ between the hygienic and unhygienic reviews.
- Unhygienic places, people talk about the ingredients of the food.

**Cuisines**
- Cleared are clearly correlated to inspection outcome even though they are not immediately related to hygiene status.
- Chinese, American, Pizza, Italian, Indian, Asian
- Hygienic: Bread, Fish, Chips, Fast Food, German, Diner, European, Sandwiches, Vegetarian

**Service & Atmosphere**
- Correlates to inspection outcome even though they are not immediately related to hygiene status.
- Unhygienic: Service, speed, seating, or wait.

**Insightful Cues**

- 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

**Features based on review text**
- Coverage: Inspection History: average penalty score, previous inspection result
- Review count & Non-positive review count

**Features Compositions**

- Meta Data
  - Cuisine
  - Zip code
  - Inspection History
- Textual Contents
  - Unigram
  - Bigram
- All Features

<table>
<thead>
<tr>
<th>Type</th>
<th>Features</th>
<th>SVM</th>
<th>MSE</th>
<th>SCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy (%)</td>
<td>66.2</td>
<td>0.23</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Inspection History</td>
<td>72.7</td>
<td>0.20</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>Unigram &amp; Bigram</td>
<td>78.4</td>
<td>0.46</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>All Features</td>
<td>81.4</td>
<td>0.19</td>
<td>0.26</td>
<td></td>
</tr>
</tbody>
</table>

*Note: MSE = Mean Squared Error, SCC = Squared Correlation Coefficients*