Data Science Institute
COLUMBIA UNIVERSITY

**Goal:** Entity-specific sentiment analysis

Our goal is to develop a tool that generates a sentiment score for individual entities in any given review. We used a random subset of 15,000 restaurant reviews from the Yelp Open Dataset to validate our model.

![Figure 1. A motivating example. Sentiments towards different products varies within a review.](image)

**Methodology**

- We train a SpaCy ER model to be able to recognize food & beverage products in reviews.
- A product list is obtained from WordNet.
- We achieved a F1 score of 91% and novelty score of 97.17%.

- **Constituency Parsing** is used to split the comment into sentences.

- Parsing Rules are used to traverse the parse tree and determine the context surrounding each entity.

- The relevant contexts are inputted into algorithms like VADER and Stanford NLP to calculate sentiment scores for each entity.

- VADER (score of 0.25) outperforms Stanford NLP (score of -0.19) using our validation technique on a baseline rule.

**Acknowledgments**

We would like to thank Felipe Penha, Sining Chen, Manoel Vilela, and Vitor Hugo for their guidance and mentorship throughout the course of the project.

---

**Validation using Yelp Star Ratings**

Validating our process is a challenge due to the lack of target labels in our dataset. To quantitatively evaluate our model, we developed a rank-based validation method that uses Yelp stars as a proxy to determine the population's sentiment toward an entity.

![Figure 3. Illustration of end-to-end validation process.](image)

Our **best rule** (rank correlation of 0.58) is as follows: For each entity, start from the minimum sentence in parse tree containing the entity. While the sentiment is neutral, replace with next shortest sentence until we reach a non-neutral sentence. (or the root).

**Example Results (Best Rule)**

**REVIEW:**

“I had 3 tacos: the Standard which was my favorite, the pork was okay and the carne was typical, no surprises. I have to say that the Prickly Pear **margarita** was the absolute best!!!!!!”

<table>
<thead>
<tr>
<th>Entity</th>
<th>Relevant Context</th>
<th>Sentiment (VADER)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pork</td>
<td>the pork was okay</td>
<td>+0.23</td>
</tr>
<tr>
<td>Margarita</td>
<td>the Prickly Pear margarita was the absolute best</td>
<td>+0.52</td>
</tr>
</tbody>
</table>

Table 1. Example of end-to-end results

**Conclusion and Future Work**

Our methodology is able to generate sentiment scores on identified entities from an arbitrary corpus, with the help of a trained ER model. These steps have been packaged as open-source software at [github.com/timjaya/neoway-brand-sentiment](https://github.com/timjaya/neoway-brand-sentiment). Next steps include comparisons with other parsing methods, such as dependency parsing.

**References**

