**Rating Restaurants from Twitter feed Vs Review Portals**

**1. Team Members**

Dharani Manian

Tanvir Bhuiyan

Sarah Garcia

**2. Problem Statement**

To find if a restaurant is associated with positive or negative attitudes based on tweets associated with the restaurant and its location and compare with other online review portals.

**2.1. Why is the problem important?**

All restaurants have a physical space that can affect a person’s mood, emotions, and experience. While restaurant goers typically rate restaurants using popular reviewing sites such as writing reviews on Google and Yelp, these ratings are not fully encompassing. Many often don’t leave reviews unless they have extreme experiences. For example, a customer leaving a very negative review over a complaint such as food poisoning.

On the other hand, users of social media sites such as Twitter constantly post updates of their daily life and activities in the form of Tweets. Our project will investigate the use of Tweets to calculate restaurant reviews attitudes and vibes, to achieve a more holistic score of how a restaurant generally makes customers feel.

**3. Research Question**

* Can we identify a restaurants’ standing (positive, negative, neutral) from the feeds of its customers in a social media network like Twitter?
* Do the sentiment analysis of twitter feeds using VADER correlate with star review ratings on review portals such as Yelp and Google?
* Can this study be used to predict the outcome of a new restaurant which does not have a handful of reviews in google or yelp?

**4. Datasets**

**4.1 Social media data**

List of restaurants and their coordinates(Latitude and Longitude) can be retrieved using Yelp Public API. Twitter Feeds associated with restaurants of Tampa Bay area are collected via Twitter API. The restaurant dataset from Yelp and Twitter data serves as input.

References:

* [*https://www.yelp.com/developers/documentation/v3/business\_search*](https://www.yelp.com/developers/documentation/v3/business_search)
* [*https://tweepy.readthedocs.io/en/latest/api.html*](https://tweepy.readthedocs.io/en/latest/api.html)
* [*https://developer.twitter.com/en/docs/tutorials/filtering-tweets-by-location*](https://developer.twitter.com/en/docs/tutorials/filtering-tweets-by-location)
* [*https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview/geo-objects*](https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/overview/geo-objects)

**5. Methodology**

**5.1.Data Collection:**

1. Creating a restaurant Dataset which consists of the restaurant’s coordinates and save them in a text file in a serializable format.

2. Using those coordinates and retrieve the geo-tagged tweets using twitter geo-objects. The primary objects are geocode(Latitude, Longitude) and Radius/bounding box which encompasses the location. The resulting tweets for each restaurant serve as another input file.

**5.2.Data Analyzing**:

3. Retweets are not collected in the 2nd step as they are not geo-tagged. Also, tweets are originated around the restaurants coordinates and hence cleaner output with minimal spams can be obtained. Relevant search tags can be used to retrieve tweets pertinent to the context.

4. Sentiment analysis is performed on each tweet collected using VADER. Output from VADER sentiment analysis will produce a polarity classification of the tweet being either a *positive, negative,* or a *neutral*opinion.

* VADER Reference - [*https://github.com/cjhutto/vaderSentiment*](https://github.com/cjhutto/vaderSentiment)

5. From the above step, majority sentiment is computed for each restaurant respectively.

6. Other review portals like Yelp will be used to collect data regarding rating of restaurants. Restaurant ratings will be classified as *positive, negative,* or *neutral*based on stars awarded. [1-3 stars – negative,3 - 3.9 =  neutral,> 3.9  = positive,  \* The numbers may vary in future after further analysis and reasoning]

**5.3. Data Visualization:**

Yelp ratings will be compared with evaluations computed from twitter sentiment analysis and study if they match or not.

**6. Related Work**

It is a subset to our first reading *“The Shortest Path to Happiness: Recommending Beautiful, Quiet, and Happy Routes in the City”*, in which sentiment is analyzed for geo locations. In contrast to the reading, Our project relies on crowdsourced tweets of people’s emotions in a restaurant space rather than urbangems.com.

**7. References**

1. Quercia, D., Schifanella, R., & Aiello, L. M. (2014, September). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proceedings of the 25th ACM conference on Hypertext and social media* (pp. 116-125).  
   [*https://dl.acm.org/doi/abs/10.1145/2631775.2631799?casa\_token=LZJ4yEOpapQAAAAA:7BbvcDRkJnWwrcWH4cfDPOccykQ\_0JR2pM-yu9zlri63Je\_hDssD0dyz9qNVP7nrAY3NdeqIoVRCEg*](https://dl.acm.org/doi/abs/10.1145/2631775.2631799?casa_token=LZJ4yEOpapQAAAAA:7BbvcDRkJnWwrcWH4cfDPOccykQ_0JR2pM-yu9zlri63Je_hDssD0dyz9qNVP7nrAY3NdeqIoVRCEg)
2. Hutto, C., Gilbert, E., (2014) VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text [*http://eegilbert.org/papers/icwsm14.vader.hutto.pdf*](http://eegilbert.org/papers/icwsm14.vader.hutto.pdf)