Elliott Goldstein

Sven Anderson

CMSC 201 Final Project Analysis

<div align="center">Final Project Analysis</div>

For my final project, I chose to analyze which kinds of foods were most well liked by New Yorkers. The geolocation of all my SearchTweets objects are the coordinates of NYC, and the radius about which I am searching is 60 kilometers. I thought it'd be an interesting topic to search because there are so many options of kinds of food available to us to choose from- so which is the most well liked? I'd decide this based on which sentiment value was the highest for a particular food. My search queries were "mexican food", "pizza", "indian food", "chinese food", and "sushi". My code collects tweets by utilizing the save method found in searchTweets. For the save method I do the following: I save my files by declaring a fileoutputstream which reads from the file given as a parameter by the client. I then add the data from the file by doing a for each loop on the tweets list where each iteration I use the writeObject method which will write a tweet onto the tweets list. Much like the save method, I do a for each loop in which I will loop over each Status object in my tweets list. I avoid adding duplicate tweets by checking the equality of the tweet's text and usernames. I add new tweets by saving them to a new file which I denote by adding the date.getTime() method to my filename. By doing this, after running my do.gather script I will have many increasingly large files and I can tell which is the largest by picking which filename is the most recent time. I created a constructor for my AnalyzeSentiment class. Creating a constructor allows me

to have a parameter filename that creates an object instance of AnalyzeSentiment. The parameter is the file for which I will populate the tweets list with. Creating this constructor allows me to have object instances of the analyzeTweets method so I don't always return the same sentiment values. If i had not made the constructor, I would have been forced to make analyzeTweets a static method which would always return the same sentiment values . I created a method loadintolist that is extremely similar to my load method except for the fact that it returns a List<Status>. I do this so I can I call it in my analyzeTweets method. However, before I analyze I must first load in my sentiments.csv file. I append from the given filename parameter given (it will be sentiments.csv in my main method) to get my RedBlackBST containing the sentiments and their value. I do this by scanning in from the file given by the client, and then I split the strings into a string array (this will separate the key and value, or word and sentiment, values). I use my Sentiment class to create a sentiment which consists of the word and it's value, and then use my Sentiment methods to get the word and sentiment value to input into my RedBlackBST, is also what I return (A RedBlackBST of words and their sentiment values). Finally, I made my analyzeTweets method.I check my list of words from my tweets against my words from my RBTree (which I use to get my words and their sentiment values). I add up my sentiments value and divide it by the length of my words array(the words array is the result of splitting each word from my tweet so it is the length of the tweet). I can analyze my different files by changing the parameter of staticload. I chose a RedBlackBST because it allows me to implement a data structure that has methods that allow me to easily get the value from inputting a

key (the get method, which was very helpful in analyze tweets) and can efficiently store lots of data thanks to how easy it is to traverse it, which is helpful because it is the data structure I'm using to store all of sentiments.csv.

<u>Analysis of My Results</u>

Sentiment Values of Different Kinds of Food in the Greater NYC Area
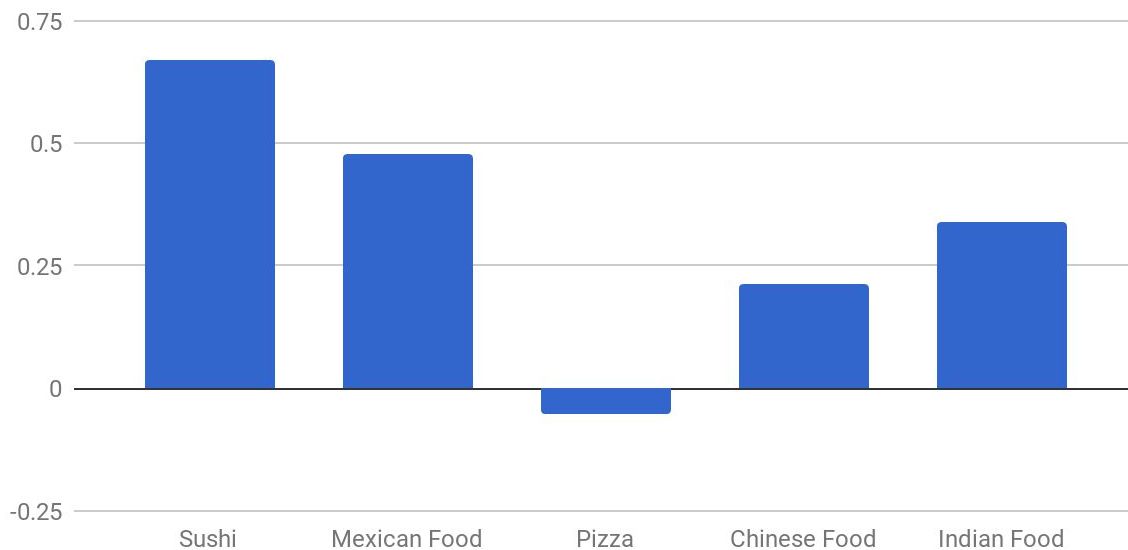
Surprisingly, pizza had by far the lowest sentiment value of any of the foods, with a -.05 average sentiment value and sushi had the highest, with roughly a .69 average sentiment value. The pizza query had the highest percentage of tweets with a negative sentiment, followed by chinese food. Pizza also had the highest average negative sentiment value of all of it's tweets that had a negative sentiment.