NLP Project: Investigating Social Value of Food Products using Population's Sentiments

Team Members: John Baldwin and Marie Pelagie Elimbi Moudio

1 Background and Project Goal

One of the great challenges our civilization is facing is developing the ability to feed our ever-growing population despite our planet's limited resources. According to reports from the world bank, a rising number of households are experiencing food insecurity due to the supply chain disruptions and reduced salaries resulting from the COVID-19 pandemic [8]. Food supply chains are becoming increasingly complex due to globalization and culture exchange [1,2]. Furthermore, the world's diet is highly dependent on only four grains: rice, wheat, corn, and soy [13]. Also, highly perishable grocery foods constitute up to 50% of all sales in retail food [9]. Supply chain managers particularly in retail groceries have a hard time managing perishable food products because of deteriorat ion, wastage, spoilage, and a short shelf life [10,11,12] . Moreso, farmers tend to rely on crop prices and profitability when deciding what crops to plant during each planting season [17]. This may seem like a straightforward decision but still requires small scale farmers to choose from a large selection of products for a limited farmland space. These challenges motivate the need to study how the community values different food items. The choice to grow produce that ends up with low future demand results in food wastage and loss of income from the farmers' perspective. As such, understanding changes in demand and supply around these key foods can inform better food policies and promote better wages for farmers and food suppliers.

Big data has been used to investigate various aspects of the food supply chain including production optimization, food quality and safety [3]. The presence of an increasing amount of available data has enabled advances in data driven insights and decision making across multiple industries including the food and nutrition sectors [4]. In recent years, the importance of text mining and analysis for food and nutrition analysis has gained more relevance and several researchers in the food supply chain sector have sought to use these text data as a decision-making tool [5]. Some researchers have explored the relationship between sentiment analysis and various food related topics, companies and reviews[5,14,15]. A study investigated public sentiment toward specific food safety incidents by developing a sentiment classification model using deep learning combined with IFoodCloud, a platform for the real-time sentiment analysis of public opinion on food safety in China [7]. Another study carried out sentiment analysis of different app based online food delivery companies using data from twitter data [16]. These studies highlight the increasing interest in understanding food services and systems using big data and social media.

Our project aims to leverage text mining and sentiment analysis tools to shed more light on what importance the community places on different foods (social value) so as to better understand if the community's sentiments towards different foods are aligned with their economic value. We plan to understand if current food pricing is aligned with public sentiment around food types. We were initially interested in three dimensions (positive vs negative, Affordable vs expensive and Healthy vs unhealthy) to get a more holistic view of people's opinions about food. We hypothesized that people tend to have strong feelings and opinions on the cost of food and where these foods lie on the healthy versus unhealthy spectrum. To do this, estimate the population's sentiments about different foods over a time period across positive and negative dimensions by building a sentiment classifier and a pre-labelled database. Our trained classifier was not sensitive enough to pick up differences between food products. We thus resolved to annotate our data by first developing an annotation guideline and ensuring that the inter-annotator agreement was greater than 0.6. Using our annotation guidelines, we then labeled 1000 coffee tweets which we used to train our model. Then we tested our trained model on 50000 tweets obtained in the past 9 years (2013-2021). We found no correlation between any sentiment (positive , negative, neutral) and the coffee prices. This analysis suggests that coffee prices are not influenced by the general public's sentiment towards the product.

2 Methods and Results

Sentiment Classifier with Supervised Learning for Positive or Negative Sentiment

We developed a sentiment classifier that we used to classify Tweets extracted from the Twitter API as positive or negative (two categories for a binary classification problem). We used the sentiment140 dataset (sentiment140.com), a dataset hosted by Stanford, as our training data for the model. The dataset consists of 1.6 million tweets that have been pre-labeled as being either positive or negative. We used BERT as our model architecture to form the actual classification mechanism for the problem.

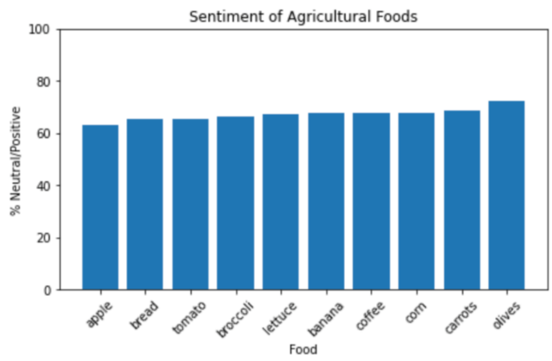
First, we prepared the data. The sentiment140 dataset only contained positive and negative labels (no neutral), so we used these two labels to train the data. We then converted each tweet into the text specifications needed to be fed into the BERT model. In this way, we restricted the tweets to a maximum of 32 tokens and added [PAD] tokens up to the 32 token limit for tweets that had less than 32 tokens. We came to choose 32 as the limit through a visual analysis of tweet length in the data. We encoded each tweet and fed it into a data loader to be used when training our model.

We trained our model using the bert-base-uncased model, and the full version of BERT (12 layers and 768 dimensions). The model was very computationally intensive to train, so we used only a fraction of our data for training (28,000 examples for training, test and development out of a dataset of 1.6 million examples), and even this took upwards of 50 minutes. We achieved an accuracy of 82.6% on our test data, with bootstrap 95% confidence intervals of 81.5% to 83.8%.

Sentiment Analysis of Tweets and Discussion

We extracted Tweets from the Twitter API using the Tweety package in order to run our trained sentiment classifier and look for differences in sentiment between foods. We specified two lists of ten keywords each. The first list contained foods that were exclusively agricultural products ('apple', 'bread', 'coffee'), whereas the second list contained foods that were more variable, including processed foods and meats such as 'ice cream' and 'steak'. We extracted 200 tweets per word, for lists totaling 2,000 tweets each. Our sentiment classifier then classified all the tweets into two categories, 'Positive/Neutral' or 'Negative'.
the dictionary-based classifiers for cost and food health.

Our results displayed in figure 1 denoting the fraction of positive (negative) classified tweets for each food type did not show major immediately visible differences in sentiment towards tweets with different kinds of foods in them. For each food category, around 65-70% of tweets were classified as neutral/positive.



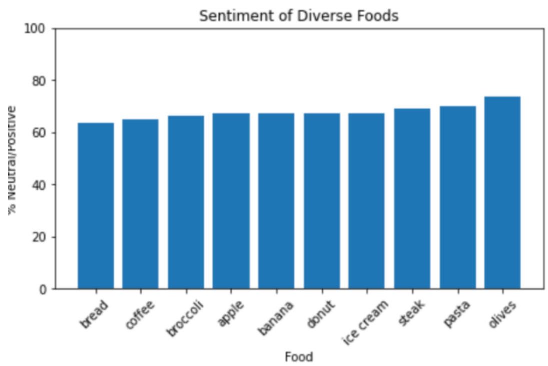


Figure 1: Fraction of positive (negative) classified tweets for each food type.

The lack of signal in distinguishing sentiment between these groups may be due to the noisiness of the underlying Twitter data. An inspection of the actual classified tweets shows many problems in the ability to extract signals from the data. For one, there is an entity matching problem; often, the chosen food represents something else than itself in the actual tweet. For instance, it was very common for tweets with the keyword 'banana' to actually be representing 'banana bread' or 'banana pudding' when looking at the tweets rather than the food commodity that is a banana. Another issue is that, while we were able to find tweets that contained the keyword for the food, the tweets containing the word themselves were often about something completely different than consuming the food, or didn't have the food as the subject of the sentence. Due to these results, we opt to label our own data so as to improve training results from our model.

Data Annotation

We created annotation guidelines specifically for sentiments (positive, negative or neutral) on tweets about food. Tweets containing the word 'coffee' but where the tweet had nothing to do with coffee were included in the neutral category as noise. Then we randomly selected a subset of tweets which were annotated separately by 2 individuals using the guidelines provided. We then created the confusion matrix and computed the observed agreement and Cohen's kappa, because the data was unbalanced in order to quantify the inter-annotator agreement.

Confusion Matrix

	Positive	Neutral	Negative
Positive	21	6	0
Neutral	10	47	0
Negative	0	1	3

Observed Agreement = 0.81
Cohen's Kappa = 0.61

Annotation-based Sentiment Analysis of Tweets

Given that our annotations had a moderate degree of similarity between annotators as measured by our Cohen's kappa score, we then annotated 1,000 tweets related to the key word 'coffee' according to these guidelines. We then trained the BERT-based model we created above on this data and observed the results. The model yielded an accuracy of 71% with 95% confidence intervals of 63% and 78%. The majority class baseline for the model was 57%, indicating that our model had a modest but statistically significant improvement of 14 percentage points over the baseline for accuracy. In addition, the F1 score was calculated to be 80% with 95% confidence intervals of 72% and 86%.

In our next step, we classified the sentiment of 50,000 tweets spanning a period of 9 years and then compared the resulting time series to coffee price data over the same time period and assessed whether there was any correlation between the two datasets. We used the snscrape command line tool to scrape tweets containing the keyword ‘coffee’ at 3-month intervals over the course of 9 years, for a total of about 50,000 tweets. We used our annotation-based classifier to classify each of these tweets as positive, neutral or negative according to the specifications made by our annotations.

We then calculated the percentage of tweets that were classified as positive sentiment over the course of 9 years for each 3-month interval we collected tweets for. We graphed the positive sentiment ratio of tweets versus coffee prices, and developed the following time series graph reflecting their relationship. Figure 2 provides a summary of our analysis investigating the correlation between prices and sentiment for coffee. We plot the trends of each sentiment (positive, negative, neutral) and coffee prices on the same axis over a period of 9 years (2013-2021). In addition we also provide correlation plots between prices and each sentiment. From visual inspection, it was tough to see if any relationship between the datasets existed in either of the graphs. To confirm our suspicions, we calculate the Pearson correlation coefficient for our variables coffee price and positive sentiment. The resulting coefficient was .11 with a p-value of .51, indicating little to no statistically significant relationship between the two datasets. Similarly we found no observable relationships between the variables and no correlation from the Pearson correlation coefficient.

Coffee prices are affected by a wide variety of factors, including weather and climate conditions, conditions in the shipping industry, and product demand. It could be postulated that sentiment towards coffee could be related to product demand which also drives coffee prices, but our above analysis shows that any connection between these variables is not relevant enough to show up in a statistically significant relationship given our data. One limitation of our analysis is that the time resolution (3-month window) is very low. Social media sentiments and events could be happening over shorter periods of time leading to undetectable effects if any. It could be interesting in the future to increase the time resolution by maybe looking at bi-weekly, monthly and even daily price and sentiment values. In addition, we only explore one food, coffee. It may be interesting to carry out a similar analysis on different food types and classes.

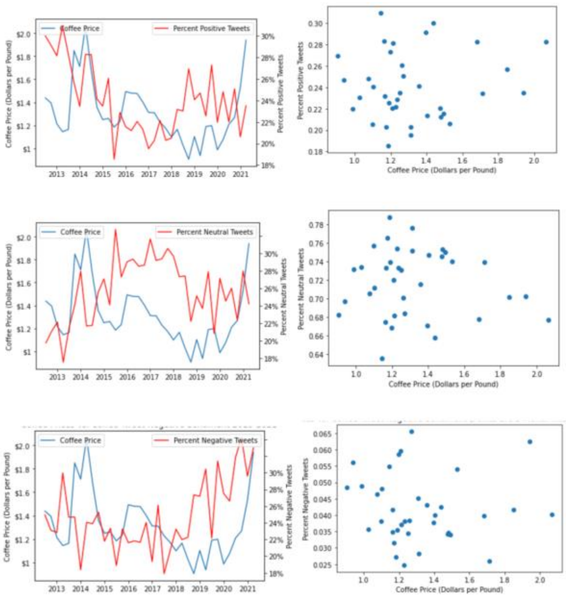


Figure 2: On the Left panel, we show the trend in coffee prices between 2013 and 2021. We also overlay our sentiment trends, positive, neutral, and negative respectively for the same period. On the right panel, from the top to bottom, we provide scatter plots of coffee prices versus sentiment, positive, neutral and negative respectively for the period of 2013-2021. We used a time resolution of 3 months to collect our data for the correlation plots.

In the future, we plan to use our sentiment analysis engine and analytic process designed above to assess the relationship between other kinds of foods and real-world data, hypothesizing relationships between different time series and then testing for any kind of relationship. In particular, we are interested in seeing if there are correlations in positive sentiment over time for the price or search term relevance of so-called fad foods, such as kale or oat milk, that received surges in popularity at various times over the past 10 years.

3 Conclusion

In conclusion, we used sentiment140 dataset, consisting of 1.6 million pre-labeled (positive, neutral or negative) tweets used for training of our sentiment classifier. We found no visible differences in sentiment towards tweets with different kinds of foods in them. In order to improve our performance, we developed annotation guidelines specifically for sentiments (positive, negative or neutral) on tweets about food. We obtained a Cohen's Kappa value of 0.61. We then created 1,000 annotations using these guidelines for the food product coffee and trained our sentiment classifier using these data. Using the resulting classifier, we classified the sentiment of 50,000 tweets with the keyword 'coffee' over the course of 9 years and assessed whether there was a relationship with coffee prices over this time period. Our results showed that there was no statistically significant relationship between these variables over the given time period. In the future, we plan to annotate other food tweets, train our sentiment classifier on these annotations and assess the relationship between food sentiment and other kinds of variables. One limitation to our model is that we only used 1000 labeled tweets to train our classifier. In the future, we would like to increase the number of labeled examples so as to improve our classifier. In addition, we will include in this looking at fad foods such as oat milk or kale to see if our food sentiment data correlates with spikes in popularity of these foods.

References

1. Christopher, M. & Lee, H. Mitigating supply chain risk through improved confidence. International journal of physical distribution & logistics management (2004).
2. Christopher, M. & Peck, H. Building the resilient supply chain. (2004)
3. Godfray, H. Charles J., et al. "Food security: the challenge of feeding 9 billion people." science 327.5967 (2010): 812-818.
4. Provost, Foster, and Tom Fawcett. "Data science and its relationship to big data and data-driven decision making." Big data 1.1 (2013): 51-59.
5. Tao, Dandan, Pengkun Yang, and Hao Feng. "Utilization of text mining as a big data analysis tool for food science and nutrition." Comprehensive reviews in food science and food safety 19.2 (2020): 875-894.
6. O'Connor, Brendan, et al. "From tweets to polls: Linking text sentiment to public opinion time series." Fourth international AAAI conference on weblogs and social media. 2010.
7. Zhang, Dachuan, et al. "IFoodCloud: A Platform for Real-time Sentiment Analysis of Public Opinion about Food Safety in China." arXiv preprint arXiv:2102.11033 (2021).
8. Food security and COVID-19. World Bank. (n.d.). Retrieved September 23, 2021, from https://www.worldbank.org/en/topic/agriculture/brief/food-security-and-covid-19.
9. J. Chung and D. Li, "The prospective impact of a multi-period pricing strategy on consumer perceptions for perishable foods," British Food Journal, 2013.
10. T. C. Aleruchi, Strategies to Minimize Perishable Food Loss in the Retail Grocery Business. PhD thesis, Walden University, 2019.
11. X. Wang and D. Li, "A dynamic product quality evaluation based pricing model for perishable food supply chains," Omega, vol. 40, no. 6, pp. 906–917, 2012.

12. R. Lukic, D. V. Kljenak, and D. Jovancevic, “Retail food waste management, “Management Research and Practice, vol. 6, no. 4, p. 23, 2014.
13. J. Woetzel, D. Pinner, and H. Samandari, “Climate risk and response,” tech. rep.,McKinsey Global Institute, 2020.
14. Ahmed, Hafiz Muhammad, et al. "Sentiment Analysis of Online Food Reviews using Big Data Analytics." *Hafiz Muhammad Ahmed, Mazhar Javed Awan, Nabeel Sabir Khan, Awais Yasin, Hafiz Muhammad Faisal Shehzad (2021) Sentiment Analysis of Online Food Reviews using Big Data Analytics. Elementary Education Online* 20.2 (2021): 827-836.
15. Sasikala, P., and L. Mary Immaculate Sheela. "Sentiment analysis of online food reviews using customer ratings." *International Journal of Pure and Applied Mathematics* 119.15 (2018): 3509-3514.
16. Trivedi, Shrawan Kumar, and Amrinder Singh. "Twitter sentiment analysis of app based online food delivery companies." Global Knowledge, Memory and Communication (2021).
17. “How Does a Farmer Decide What to Plant?” *Good in Every Grain*, 5 Feb. 2020, https://goodineverygrain.ca/2015/04/28/how-does-a-farmer-decide-what-to-plant/.
18. SAGE Research Methods. Learn about Dictionary-Based Sentiment Analysis in Python with Data from the Economic News Article Tone Dataset (2016), Sage Publications, 2019, https://methods.sagepub.com/dataset/howtoguide/dictionary-based-sentiment-analysis-in-econnews-2016-python.
19. Silva, Nadia Felix F. Da, Luiz FS Coletta, and Eduardo R. Hruschka. "A survey and comparative study of tweet sentiment analysis via semi-supervised learning." *ACM Computing Surveys (CSUR)* 49.1 (2016): 1-26.