# Understanding User Engagement: A Sentiment Analysis of Food Security Organizations' Social Media


Janet Flores[1], Dr. Marcelo Sztainberg[1], Dr. Nancy Wrinkle[1], Dr. Benito Mendoza[2]

Northeastern Illinois University, Chicago, IL 60625[1]

New York City College of Technology, Brooklyn, NY 11201[2]



## Abstract

Food insecurity, the lack of access to enough nutritious food for a healthy lifestyle, is a problem many in the U.S. are facing, in particular minorities (17.2% Hispanic households, 2020). Food security organizations are using social media to communicate the services they provide to food-insecure individuals but have been ineffective at reaching enough people. The goal of this project is to improve the connection between food security organizations and Hispanics in an effort to expand the use of services amongst them. This will be achieved using sentiment analysis, the process of identifying the emotion behind a piece of text. Focusing on the use of language by organizations on their social media will yield insight on how the language being used can increase user engagement. We work on the Facebook pages of three different organizations located in Illinois, implementing four different algorithms: VADER, Multinomial Naïve Bayes, Random Forest, and Multinomial Logistic Regression. A curated dataset (Twitter Airline Sentiment) is used to train these models. The results of the sentiment analysis are combined, assigning a final sentiment to each of the Facebook posts.


## Introduction

Despite its status and wealth, the U.S. currently struggles to feed all its people. In 2020 alone, there were 13.8 million food insecure households. Aside from money being an issue, it seems many Hispanics simply failed to take advantage of the services, like food pantries, that were made available to help them. This stemmed from limited access and usage of information systems, which itself was a result of a variety of different factors. With food security organizations primarily making use of social media to spread word among their communities about the resources they offer, those unfamiliar or with no access are unable to partake. To address this issue, sentiment analysis is performed on the Facebook pages of various food security organizations. In doing so, we will have a deeper understanding of how to better inform people about access to the services.

## Background

Sentiment analysis, also known as opinion mining, is the process of identifying the emotion expressed in a piece of text. This technique is often used in products and services reviews. Businesses perform sentiment analysis on customer feedback to monitor product sentiment and better understand customer needs. In healthcare, sentiment analysis is used to improve healthcare quality [1]. When multiple machine learning algorithms are used to perform the sentiment analysis, ensemble methods can be used to improve the predictive performance. One of the simplest methods used is voting, where the prediction with the majority votes is selected [6].

## Methodology/Approach

- The Facebook pages of the following organizations were analyzed:

| Organization | Type | Followers |
|---|---|---|
| Greater Chicago Food Depository | Charity Organization | 40K |
| Northern Illinois Food Bank | Charity Organization | 26K |
| Central Illinois Foodbank | Charity Organization | 5.9K |

- Multi-class classification: positive, negative, neutral
- Models: VADER, Multinomial Naïve Bayes (MNB), Random Forest (RF), Multinomial Logistic Regression (MLR)
- VADER (Valence Aware Dictionary and sEntiment Reasoner) is a pre-trained lexicon-based model that is sensitive to both polarity and intensity. Since VADER is optimized for social media data, it can handle: emojis, emoticons, slang, acronyms, amplifiers.
- VADER outputs four scores: positive, negative, neutral, compound. It has the following threshold values for classifying text:

  **Positive Sentiment:** compound score $\geq 0.05$

  **Neutral Sentiment:** $-0.05 <$ compound score $< 0.05$

  **Negative Sentiment:** compound score $\leq -0.05$
- MNB, RF, and MLR are all supervised machine learning (ML) algorithms. ML models require preprocessing and training to perform sentiment analysis
- Preprocessing: tokenization, lemmatization, stopword removal, replacing emojis with name of the emoji
- The Twitter Airline Sentiment Dataset (14,640 tweets) was used to train and test the ML models
- To train the ML algorithms, the text data must be converted into numerical feature vectors. This is achieved using the Bag-of-Words (BoW) model. BoW converts text into vectors by counting the number of times a word appears, as shown in Figure 1. The features in the model are distinct words from the text data, and the feature values are the word counts.

| text | chicago | communication | different | flight | get | late | leave | minute | thank |
|---|---|---|---|---|---|---|---|---|---|
| thank get different flight chicago | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| leave minute late flight communication… | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |

*Figure 1: Bag-of-words of Twitter data*

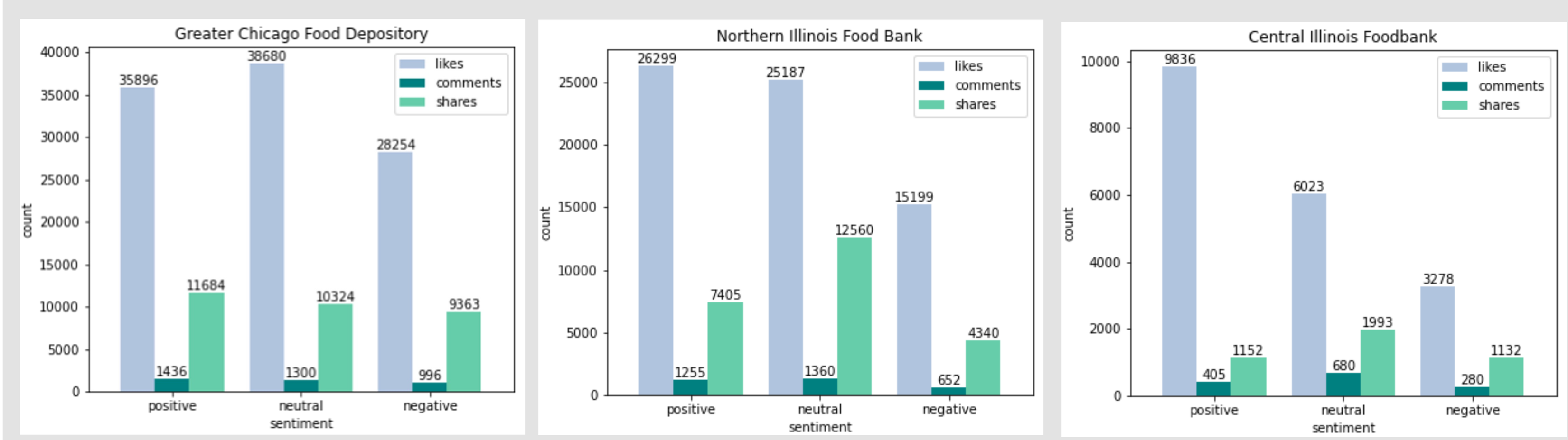
- Twitter dataset was imbalanced with 62.7% negative tweets, 21.2% neutral tweets, and 16.1% positive tweets. The resampling technique Random Oversampler was used to improve ML models' performances.

## Results/Discussion

- Neutral sentiment is the one expressed the most by all three organizations
- For Greater Chicago Food Depository, negative posts are more prominent than positive posts
- Three types of user engagement selected for analysis: likes, comments, shares
- Users engage with posts using the "Like" button most commonly

| Organization | Number of Posts | Positive | Neutral | Negative |
|---|---|---|---|---|
| Greater Chicago Food Depository | 3,480 | 29.7% | 38.3% | 32% |
| Northern Illinois Food Bank | 3,605 | 38% | 38.6% | 23.4% |
| Central Illinois Foodbank | 1,966 | 38.8% | 42% | 19.2% |

*Figure 4: Posts classified positive, negative, neutral*




*(a): Greater Chicago Food Depository* *(b): Central Illinois Foodbank* *(c): Central Illinois Foodbank*

*Figure 5: Likes, comments, and shares each organization received on their posts*

- Word clouds display different words in varying sizes. The size is related to the frequency of use, with the most frequently used words appearing larger.
- Word clouds were generated for each class (positive, negative, neutral) per organization. This allowed us to identify the words that have a positive, negative, or neutral connotation, along with their level of impact.
- It allowed us to identify what words are most commonly used across all three food security organizations. "Thank" is one such word. From Figure 6a we can see that within the Greater Chicago Food Depository, "Thank" is one of the most commonly used word with a positive connotation. With regards to both the Northern Illinois Food Bank and the Central Illinois Foodbank, "Thank" is the most commonly used word with a positive connotation.
- A word may have a different connotation within each organization. Figures 6a, 6b, and 6c show how differently each organization uses the word "help".


*(a): Positive posts made by Greater Chicago Food Depository*

*(b): Negative posts made by Central Illinois Foodbank*

*(c): Neutral posts made by Northern Illinois Food Bank*

*Figure 6: Word clouds generated from each organization's posts*

## Methodology/Approach (Cont'd)

- Accuracy scores of each model:

| | VADER | MNB | RF | MLG |
|---|---|---|---|---|
| Accuracy | 54.1% | 74.0% | 74.5% | 74.2% |

*Figure 2: Accuracy scores of the models*

- Sentiments: positive = 1, neutral = 0, negative = -1
- Each Facebook post was assigned a sentiment (1, 0, -1) by each model

| post_text | vader_sentiment | RF_sentiment | MNB_sentiment | MLR_sentiment |
|---|---|---|---|---|
| Give today so more kids have the ba… | 1 | 0 | -1 | 0 |
| We are always impressed by the amo… | 1 | 1 | 1 | 1 |
| Hundreds in our community are facin… | -1 | -1 | 1 | 1 |

*Figure 3: Sentiment analysis performed on text using four models*

- We wanted to assign a final sentiment to each post. A voting ensemble would not work because there were instances where there was a tie, as is shown in the last post in Figure 3. Instead, we use the weighted sums of the sentiments, which are calculated using the accuracy scores of the models
- Weighted sums: $w_{positive}$, $w_{neutral}$, $w_{negative}$
- If a model with an accuracy of $x$ classifies a post as $i$, $x$ is added to $w_i$
- The last post in Figure 3 was classified negative by VADER and MNB, and positive by RF and MLR. Using the accuracy scores shown in Figure 2, the final sentiment assigned to the post (positive) was selected as follows:

$$w_{positive} = 74.0\ (\text{MNB}) + 74.2\ (\text{MLR}) = 148.2$$

$$w_{negative} = 54.1\ (\text{VADER}) + 74.5\ (\text{RF}) = 128.6,\quad w_{neutral} = 0$$

## Conclusion

- A combination of lexicon-based models and ML models were used to perform a sentiment analysis on Facebook posts. An ensemble technique was then used to improve the sentiment analysis.
- We identified words with positive connotations, words with negative connotations, and words with neutral connotations. For each class (positive, negative, neutral), we were able to identify the words that have a larger impact
- Words with positive connotations should be used by food security organizations. Words with negative connotations should be avoided.
- Users are more likely to engage with posts that convey positive or neutral emotions

## Future Work

- Analyze social media pages that use emojis regularly to measure the impact emojis have on the sentiment of their posts
- Obtain demographics of users who visit/interact with the social media pages of food security organizations


## Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants number 2131291. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

I would like to express my most sincere gratitude to my mentors, Dr. Marcelo Sztainberg, Dr. Benito Mendoza, and Dr. Nancy Wrinkle, for their guidance and support. I would also like to thank fellow research student Elizabeth Soto for collaborating with me on this project. Finally, I would like to thank CAHSI for sponsoring the student's trip to GMiS 2022.


## References