14th National Convention on Statistics (NCS)
Crowne Plaza Manila Galleria, Quezon City
October 1-3, 2019

AN ALTERNATIVE METHOD TO ESTIMATE CONSUMER SATISFACTION USING SOCIAL MEDIA DATA: THE CASE OF THE DEPARTMENT OF FOREIGN AFFAIRS

by

Angelyn K. Mananghaya and Zita V.J. Albacea

For additional information, please contact:

Author's name	Angelyn K. Mananghaya
Designation	Instructor
Author's name	Zita V. J. Albacea
Designation	Professor
Affiliation	University of the Philippines Los Baños
Address	Institute of Statistics, Physical Sciences Building, University of the Philippines Los Baños, College, Laguna, Philippines 4031
Tel. no.	(049) 536-2381
E-mail	akmananghaya@up.edu.ph, zvalbacea@up.edu.ph

AN ALTERNATIVE METHOD TO ESTIMATE CONSUMER SATISFACTION USING SOCIAL MEDIA DATA: THE CASE OF THE DEPARTMENT OF FOREIGN AFFAIRS

by

Angelyn K. Mananghaya and Zita V.J. Albacea[1]

ABSTRACT

Collecting feedback from customers is instrumental for the improvement of services offered by an agency. In this study, social media data were used to develop a way to estimate consumer satisfaction on a government service, which can be a faster and cheaper alternative to traditional consumer satisfaction surveys but can also generate reliable estimates. Twitter® data related to the Department of Foreign Affair's official account (@DFAPHL) were used for the study. A lexicon-based sentiment analysis, using English and Tagalog dictionaries, was used to determine the sentiment score of all tweets, which were then classified into positive, neutral, and negative. Additional results showed that sentiment score of a tweet has a very weak association with a tweet's number of words, the day and time it was sent, and its language. To determine the proportion of users with a positive satisfaction to the agency, the initial dataset of tweets was filtered such that only the most recent tweet of a user is included. Results showed that around 4 out of every 10 users showed a positive satisfaction based on the sentiment of his tweet. The statistical properties of the estimated proportion were then evaluated using bootstrap resampling, where results showed that the estimate is accurate, precise, and consistent.

1. Introduction

Government services are comparable to a citizen's human rights and should be accessible and available to all. Although one of the goals of the government is to provide excellent service to the people, leaders recognize that citizens are not often satisfied with the services offered by the government. Some agencies may focus more on completing tasks and accommodating more people rather than delivering a satisfactory experience for the citizens (Baig et al., 2014).

The best way to identify the failing points of an agency is to solicit feedback from the clients. By measuring the customers' satisfaction regularly, agencies will be able to identify the specific services they are dissatisfied with, the degree of their dissatisfaction, and what improvements can be done across agencies to raise the overall satisfaction of clients. Enhancing the customer experience is crucial to build public trust and promote integrity of the government agencies.

A way to collect customer feedback is by conducting surveys on a periodic basis. However, a drawback of this method is that the agencies receive the feedback after a certain period and may not be relevant to the current issues anymore (Mullins, n.d.). Customer satisfaction is best tracked in real-time which cannot be achieved by using traditional data sources and surveys. Thus, more diverse measuring methods and alternatives are being sought to produce real-time information.

One identified alternative is to use Big Data. People have started to shift from traditional communication tools to internet services due to easy accessibility. In the Philippines alone, around 67 million Filipinos, or 63% of the total population, are internet users, which makes the Philippines rank 12th among the top internet users in the world (Porcalla, 2018).

[1] Instructor and Professor, respectively, Institute of Statistics, University of the Philippines Los Baños

Twitter® (or Twitter, which will be used for the rest of this paper), one of the widely-used social media sites, generates over 500 million posts per day from over 300 million active users. Users in Twitter can express their thoughts on a limitless number of topics by sending tweets which have a maximum of 280 characters. Users can communicate with other users, share their mood, experiences, and daily activities, follow official accounts for latest news, or contact accounts from companies or government services to ask for help or leave feedback. All this massive amount of data is being generated real-time, and thus becoming a popular data source for research. Quite a lot of research papers that made use of Twitter data are focused on sentiment analysis.

Sentiment analysis, or opinion mining, is a way of identifying the opinion and categorizing the polarity of a given text, which could be positive, negative, or neutral, by looking at the content and structure of the text. This analysis is useful to detect customers' satisfaction in company products, movie reviews, and opinions on political news. Consumers can also use this analysis to do research on a product or service they want to avail. By using sentiment classification techniques on the data obtained from Twitter, businesses and organizations can identify the overall mood and satisfaction of their clients to gather critical feedback, and therefore implement interventions more promptly.

To keep up with the trend and to take advantage that an increasing number of people who use the internet, government services also have started to utilize social media, specifically Twitter, by creating official accounts in which people can access publicly. In addition to being a source of news and announcement, these official accounts can also answer queries and gather feedback from people. Since the tweets to these accounts can be publicly accessed, it will be interesting and meaningful to gather these tweets to conduct an analysis to determine their characteristics and sentiment which can be linked to a customer's satisfaction with the service.

One government service with an active Twitter account is the Department of Foreign Affairs (DFA). Traveling abroad is becoming more popular, that is why more people are getting their passports done or renewed at the DFA. Twitter users can choose to contact the DFA official Twitter account instead of calling or going to the DFA office. These tweets can therefore be collected and used to know the sentiments of the DFA clients on the services being offered.

By making use of information from their Twitter account, DFA will be able to track a portion of their consumers in real-time. Government agencies like DFA will be able to respond to their consumers more promptly and plan changes accordingly. This can also be a cheaper and faster alternative to having a customer satisfaction survey. Also, this may serve as a catalyst for government agencies to improve their social media presence and utilize other social networking platforms for their services.

This paper intends to present the results of a study which aimed to develop a methodology to determine the satisfaction of users on a government service, in particular the DFA services, using data collected from Twitter.

2. Methodology

2.1 Data Source

The data used in the study were the tweets obtained from the DFA's official Twitter Account (https://twitter.com/DFAPHL). All publicly posted tweets that mention @DFAPHL were collected via the Twitter Application Programming Interface (API) using the 'rtweet' package, which was developed by Kearney (2018), in R from November 4, 2018 to February 4, 2019. Data cleaning and pre-processing were done to exclude spam or duplicate tweets. Tweets identified as another language other than English or Tagalog were also removed from the dataset.

2.2 Data Analysis

Descriptive statistics and graphs were generated to summarize the characteristics of the tweets. Word clouds were constructed to determine which words appear frequently among all tweets. In addition, measures of association were used to determine if the sentiment of a tweet is related to its characteristics. Some characteristics include the time and date the tweet was posted, language of the tweet, and the number of words in the tweet.

For the sentiment analysis, the words from each tweet were gathered and compared to a dictionary containing words and their corresponding polarity. The list of positive and negative English words used is from the paper of Hu and Liu in 2004. The list of positive and negative Tagalog words is from Chen and Skiena in 2014. To have a better Tagalog dictionary for the analysis, the researcher inspected the dictionary to remove words that may have been incorrectly classified. Further, all words from the Tagalog tweets were extracted, classified based on sentiment according to the researcher’s judgment, and were added to the Tagalog dictionary.

To get the sentiment score of a tweet, positive and negative words were tagged with a +1 and -1, respectively. Words that are not in the dictionaries were tagged with a ‘0’ and were considered neutral words. The overall sentiment of the tweet is acquired by adding the polarity scores of all words in a tweet.

Afterwards, the tweets were categorized into positive, neutral, or negative based on the computed sentiment scores to obtain the proportion of tweets with a positive sentiment. Also, to obtain the proportion of users with a positive satisfaction based on his tweet, the tweets were initially filtered such that a user only has one tweet in the dataset. Only the most recent tweet of a user was retained in the dataset.

Bootstrap resampling was done to come up with an estimate of the proportion of Twitter users that show a positive satisfaction to DFA based on his tweet and generate its distribution. Resampling was done at different sample sizes to observe the behavior of the proportion as the sample size increased. The tweets from the unique users was used as the pseudo-population in the analysis. In addition, the statistical properties of the estimate were evaluated by determining the bias and standard error of the estimates.

3. Results and Discussion

3.1 Characteristics of the Tweets

A total of 4,503 tweets was collected. Only the set of tweets that were coded as English or Tagalog was used in the analysis, while the rest was omitted from the data set. After preliminary data processing, only 1,425 tweets were retained and used for the succeeding analyses.

Figure 1 shows the total number of tweets per day from November 4, 2018 to February 4, 2019. There is an average of 15 tweets in a day for the whole time period. The number of tweets in a day deviates from the mean by 10.7228. The minimum number of tweets is 1 on both December 23 and 25, 2018 and the maximum is 56 tweets on January 10, 2019. Upon inspection, there is no significant event that happened on January 10, 2019 that could have led to this high frequency of tweets. Although it could be noted that this day is within the first fifteen days of the year and people had already returned to their usual activities during this time.

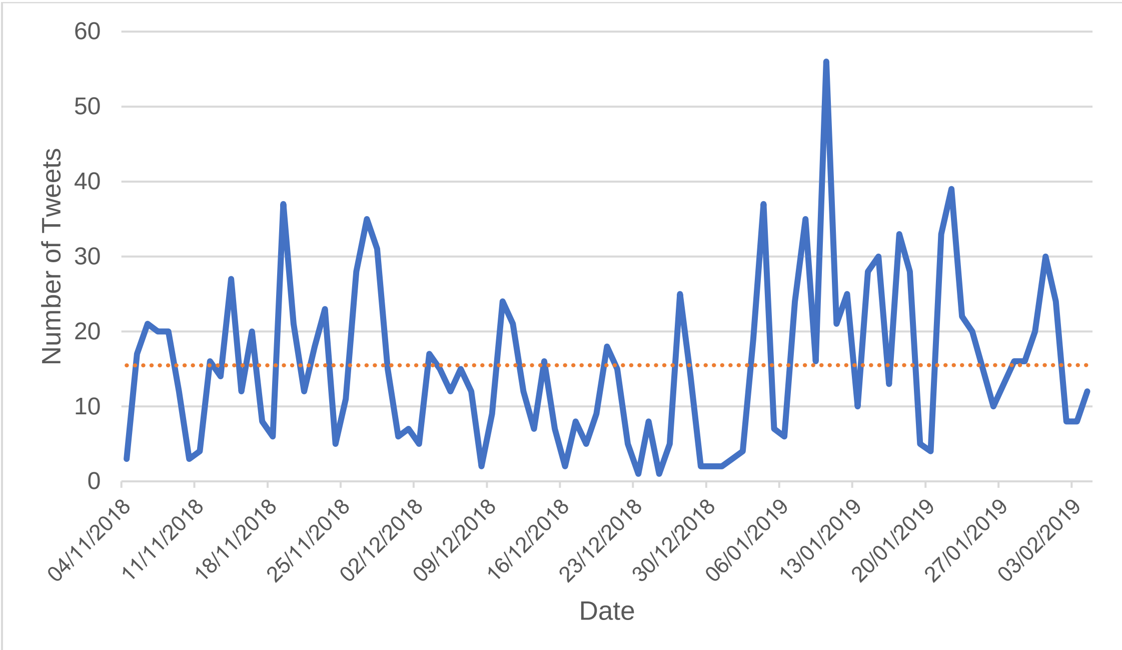


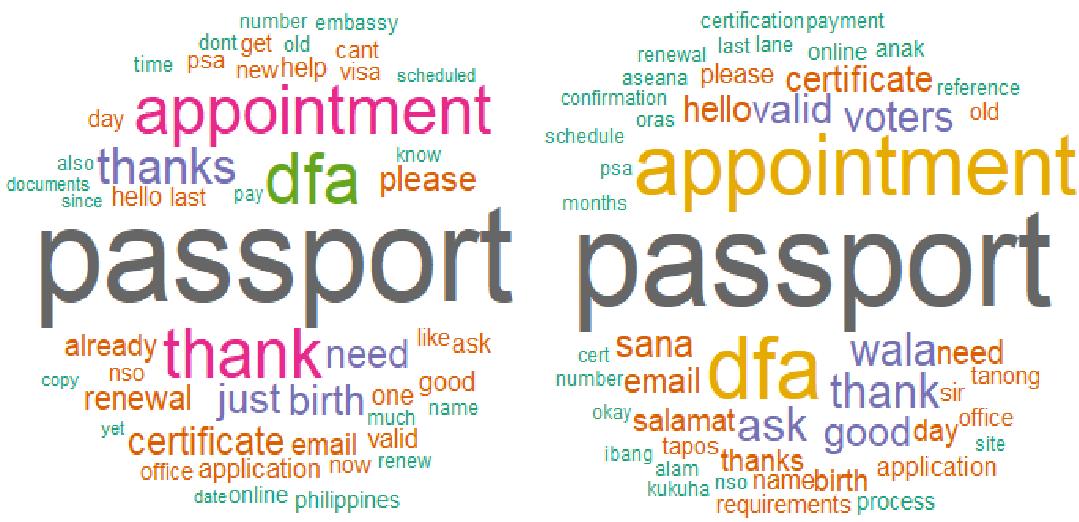
Figure 1. Total number of tweets to @DFAPHL per day from November 4, 2018 to February 4, 2019.

Additional results show that tweets are least frequent on the weekends. This is expected since the regular operating days of DFA are weekdays only, so users are less likely to contact the agency on a weekend. The tweets are most frequent on a Thursday. Users might be rushing on sending out tweets before the week ends so that actions can be done before the weekend.

Majority of the tweets are sent in the morning (12:00AM to 11:59AM), with the most frequent tweets occurring at 3:00AM to 5:59AM. In addition, only 445 (31.23%) of all tweets are sent within 8:00AM to 5:00PM, which is the working hours of the DFA offices. A reason why users are more likely to tweet earlier in a day is that there might be a higher chance of getting noticed by the agency once the office hours start at 8:00 in the morning. There is also a possibility that users send their tweets during early morning when the internet connection is much faster and there is fewer exchanges of messages at this time of the day.

English tweets are more frequent then tweets sent in Tagalog. This is probably because the DFA twitter account only posts tweets in English, so users tend to send tweets in English as well. Also, DFA is a government office, so users tend to be more formal in sending their tweets.

To determine which words appeared the most frequent among the tweets, word clouds were created by using the ‘wordcloud’ package in R developed by Fellows in 2018. As seen in Figure 2, the most frequent words for both English and Tagalog tweets is ‘passport’ and ‘appointment’. This result is expected since the main service offered by DFA is their passport processing service. Also, an online appointment is required before one can process their passport, which is why ‘appointment’ is also one of the most frequent words. In addition, some words with a positive emotion are also present in the word cloud, such as ‘thank’, ‘thanks’, and ‘good’. Its Tagalog counterpart ‘salamat’ is also present in the word cloud for Tagalog.



English Tagalog

Figure 2. Word cloud of the tweets by language.

3.2 Sentiment of All Tweets

Before starting the sentiment analysis, the Tagalog word dictionary written by Chen and Skiena (2014) was initially reviewed to check if the dictionary is usable for the analysis. The initial Tagalog dictionary contains 812 positive and 1046 negative words. Despite being labeled as Tagalog dictionary, the list contains both English and Tagalog words. The list was compared to the English dictionary by Hu and Liu (2004) and duplicate English words were removed. Also, words that were incorrectly classified were removed from the list.

To further expand the dictionary and strengthen the accuracy of the word-matching, all words from the tweets labeled as Tagalog were extracted. A total of 1,933 extracted words were obtained. Upon investigation, the words are a combination of English and Tagalog words. There are no other Filipino languages present other than Tagalog. The extracted words were then classified and added to the dictionary. The final Tagalog dictionary used for the analysis contains 588 positive and 859 negative words. The combined English and Tagalog dictionary has a total of 2,597 positive and 5,642 negative words

First, the number of positive words per tweet were counted. The minimum and maximum number of positive words per tweet are 0 and 6, respectively. On the average, the number of positive words per tweet deviates from the mean of 0.7705 by 0.9518. Fifty percent of the tweets have at most one positive word. There are only 3 out of 1425 tweets (0.21%) with 6 positive words. Also, the mode is 0, which indicates that most of the tweets have no positive words.

The number of negative words per tweet were then counted. Majority (64.49%) of the tweets have no negative words. The minimum and maximum number of negative words are 0 and 6, respectively. On the average, the number of negative words in a tweet deviates from the mean of 0.5326 by 0.8871. In addition, there are few extremely high number of negative words in a tweet. Though the shape of the distribution of negative words is similar to that of positive words, there are more tweets with no negative words than tweets with no positive words. This means that overall, there are more positive words among all tweets than negative words.

The overall sentiment scores of each tweet were then obtained by counting the occurrences of positive (+1) and negative (-1) words in a tweet. Among all the tweets, the minimum sentiment score is -6, while the maximum score is 6. Most of the tweets have a sentiment score of 0. On the average, the sentiment score of the tweets deviates from the mean of 0.2379 by 1.2978. Also, there are few extremely high sentiment scores as indicated by the positive measure of skewness (SK=0.5499) and the shape of the distribution in Figure 3. In addition, 50% of the tweets have a sentiment score of at most 0.

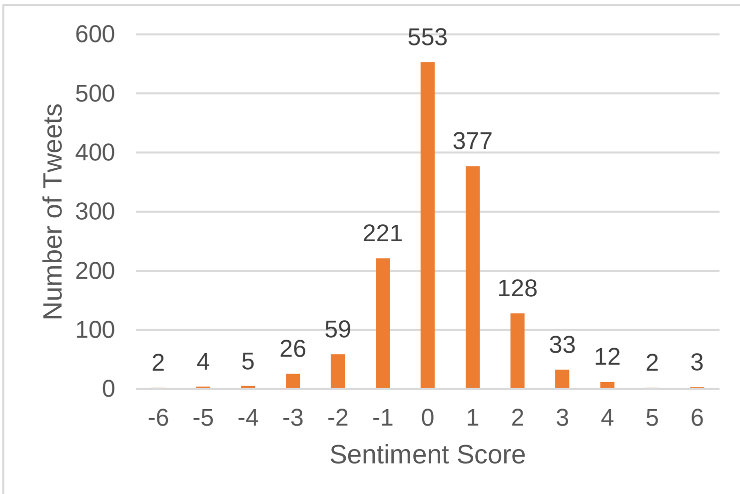


Figure 3. Distribution of the sentiment scores of the tweets.

After obtaining the sentiment scores, each tweet was then categorized into positive, neutral, or negative according to the corresponding sign of the sentiment score. Table 1 shows the distribution of the categorized sentiments. The proportion of the tweets with a positive sentiment is 0.3895 which means that around 4 out of every 10 tweets have a positive sentiment, on average. Even if majority of the tweets do not have a positive sentiment, it is worth noting that there are more tweets with a positive sentiment than negative sentiment. This means that users gave more positive tweets or feedbacks than negative tweets to the agency.

Table 1. Distribution of the sentiment categories of the tweets.

SENTIMENT CATEGORY	FREQUENCY	%
Positive	555	38.95
Neutral	553	38.81
Negative	317	22.25
Total	1425	100.00

3.3 Association of Sentiment Score with the Tweet Characteristics

The total number of words in a tweet has a negative very weak linear relationship with the sentiment score of a tweet (r = –0.1418). This means that as the word count of a tweet increases, the sentiment score of a tweet decreases, and vice versa. Similarly, the language (coded as English=0; Tagalog=1) and time in a day (coded as AM=0; PM=1) a tweet was posted showed a negative association with the sentiment score (r_{pb} = –0.1258 and r_{pb} = –0.0522, respectively) . This means that the sentiment score of a tweet tends to be higher if the language used is English, and if the tweets was posted in the morning. Furthermore, only 0.21% of the variation in the sentiment scores can be accounted for by the day of the week the tweet was posted. This shows that there is a very weak association between sentiment score and the which day in a week the tweet was posted. All values of the association measures indicate a very weak association between sentiment score and the four variables, which means that the changes in sentiment score given a change in tweet characteristic are not apparent, and vice versa.

3.4 Estimation of Satisfaction of Users Based on Tweets

To estimate the proportion of users with a positive tweet, the tweets were checked to see if a user only tweeted once during the three-month period of data collection. If a user tweeted more than once, only his most recent tweet was retained in the data set. From all the 1,425 tweets, only 781 unique users were identified. A user tweeted approximately twice during the data collection period, on average. Most users tweeted only once during the three-month span, and there is one user who tweeted 21 times, which is the maximum value. There are few users with an extremely high number of tweets as indicated by the positive measure of skewness (SK=1.5160).

The sentiment scores of each tweet were evaluated again using the updated data set where only one tweet is recorded per user. The minimum sentiment score is -5, while the maximum score is 6. Most of the users have a tweet with a neutral emotion, as indicated by the zero value of sentiment score. On average, the sentiment score of a user's tweet is 0.2945, which is positive in value but still close to 0 (neutral category). Furthermore, there are few users with an extremely high sentiment score as indicated by the measure of skewness (SK=0.6796). In addition, 50% of the users have a sentiment score of at most 0, indicating a neutral emotion.

After categorizing the sentiment scores, it can be seen in Table 2 that the proportion of users who are satisfied with the services of DFA is 0.4135, as indicated by the tweets that were classified as positive, with a standard error of 0.0176. This estimate is reliable with a coefficient of variation of 4.3%. Majority of the users (58.65%) do not show positive satisfaction, as indicated by their tweets that were classified as either neutral or negative. The estimate shows that around 4 out of every 10 users show a positive satisfaction based on his tweet. Also, looking at only the positive and negative categories, the number of users that show a positive satisfaction are almost twice as those who show a negative satisfaction, which indicates a good sign to the agency.

Table 2. Distribution of the sentiment categories of the tweets from unique users.

SENTIMENT CATEGORY	FREQUENCY	%
Positive	323	41.35
Neutral	289	37.00
Negative	169	21.64
Total	781	100.00

3.5 Bootstrap Resampling

Using bootstrap resampling at different values of sample size with B=1000, the proportion of users with positive satisfaction range from 0.4128 to 0.4136. The range of values per sample size decreases as the sample size is increased. Similarly, the standard errors also become smaller with an increase of the sample size. These indicate that the dispersion of the proportion values decreases as the sample size increases.

The measures of skewness do not have the same signs for all sample sizes, but the values are relatively close to 0, which indicates that the distribution of the proportion of users with positive satisfaction are somewhat symmetric. The shape of the distributions can also be seen in the boxplots in Figure 4.

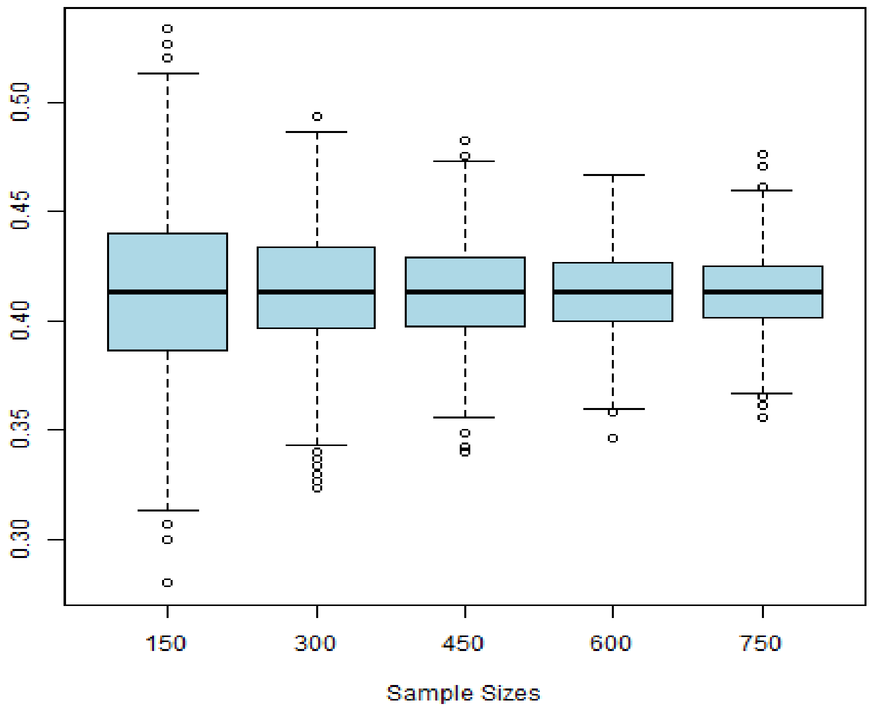


Figure 4. Distribution of the proportion of users with a positive satisfaction at different sample sizes with bootstrap resample of B=1000.

Likewise, Table 3 shows some measures to assess the statistical properties of the estimated proportion of users with a positive satisfaction. The values of the bias are relatively close to zero. This means that the estimated proportion of users with a positive satisfaction is empirically unbiased in predicting the actual proportion of users with a positive satisfaction. Also, the low values of the standard errors indicate that the estimates are also have high precision. In relation to the sample size, it can be noted that the biggest absolute decrease in bias occurred when sample size was increased to 300. Likewise, the notable decrease in standard error is when sample size was increased to 300. In general, the values of the bias and standard error both decrease as sample size increases, which indicates and supports that having a higher sample size is more ideal when doing the analysis.

Table 3. Measures of the bias and standard error for the estimates of the proportion of users with a positive satisfaction at different sample sizes with bootstrap resample of *B*=1000.

B	SAMPLE SIZE	BIAS	STANDARD ERROR
1000	150	-7.6568×10^{-4}	0.0410
	300	-4.8234×10^{-4}	0.0288
	450	-3.4345×10^{-4}	0.0233
	600	5.9323×10^{-5}	0.0198
	750	4.8990×10^{-5}	0.0177

Furthermore, the consistency of the estimate was evaluated by checking the behavior of the means and variances at different sample sizes. The previous statistics showed that the estimate of the proportion of users with a positive satisfaction approaches the true value as sample size increases. Likewise, the variances also approach zero as the sample size increases. Therefore, the estimate of the proportion of users with a positive satisfaction is said to be consistent.

3.6 Summary of the Proposed Methodology for Estimation

The general methodology that was used to obtain the proportion of users with a positive tweet were summarized in Figure 5. One must initially gather the data from a selected social networking site, in this case, Twitter, by specifying the account used by the agency to communicate with their customers. Afterwards, data preprocessing should be done to exclude irrelevant tweets from the dataset. The retained tweets will then be cleaned to prepare for analysis.

Before doing the analysis, the dictionaries of English and Tagalog words can also be updated. There may be words that differ in context depending on the nature of the agency, so these can be incorporated in the existing dictionaries for a better analysis.

In doing the sentiment analysis, one can opt to examine first all the tweets included in the dataset. This is to have an overview on the sentiment scores for all tweets received by the agency. Then, to obtain the proportion of users with a positive satisfaction, the tweets must be filtered such that a user only has one tweet in the dataset. Afterwards, the sentiment scores of the tweets will be obtained and categorized into positive, neutral, and negative which will then be used to compute for the estimate of the proportion.

Moreover, the agency can also investigate the tweets tagged as negative to see the specific concerns of the users that may have led the user into tweeting negatively. They can also investigate the tweets tagged as positive to see which of their policies or services are favored by the users.

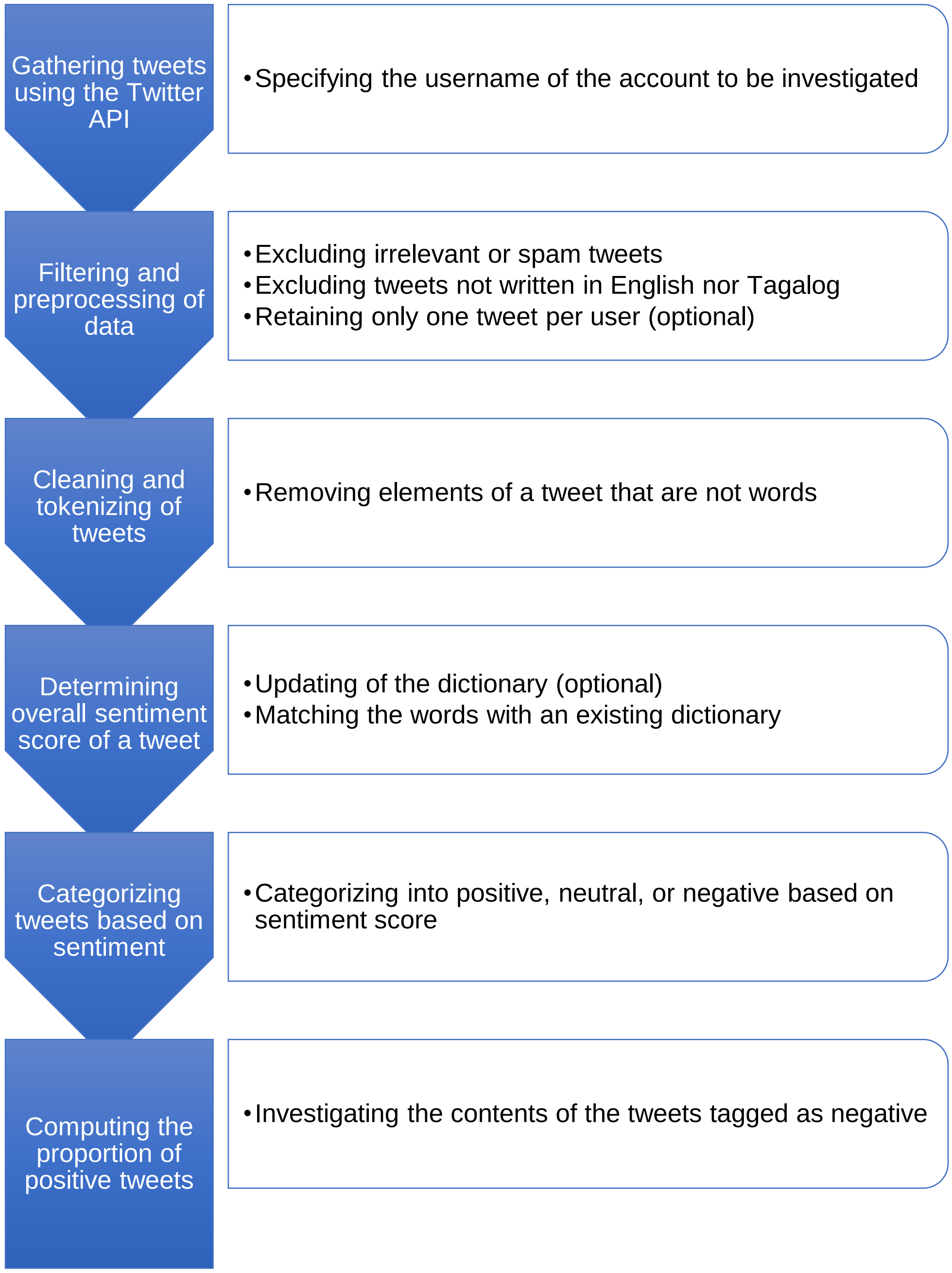


Figure 5. Flowchart of the methodology to conduct the analysis.

4. Concluding Remarks

The study made use of social media data, specifically Twitter, to develop a way to estimate consumer satisfaction of users on the services offered by the Department of Foreign Affairs (DFA). Publicly posted tweets that mention the official account of DFA (@DFAPHL) were collected from November 4, 2018 until February 4, 2019 using the Twitter API.

The consumer satisfaction of a Twitter user was determined by focusing on the sentiment polarity of a given tweet. A lexicon-based sentiment analysis was used to determine the sentiment score of a tweet, which was then classified into negative, neutral, or positive. The analysis was initially run using all the tweets. The mean sentiment score of all tweets was 0.2379 and the proportion of positive tweets is 0.3895.

To determine the proportion of users with a positive sentiment score, the tweets were again filtered to remove multiple tweets by a user and retain only the most recent tweet of a user. Out of the 781 users, the mean sentiment score is 0.2384 and proportion of users with positive tweets is around 0.4315. This means that around 4 out 10 Twitter users are show positive satisfaction to the services offered by DFA.

To assess the statistical properties of the estimated proportion of users with a positive satisfaction based on his tweet, bootstrap resampling technique was implemented using different sample sizes with the same number of bootstrap resamples of 1000, having the 781 tweets from unique users as the pseudo-population. Results showed that the proportions at different sample sizes are close to the true value and all have low measure of standard error. This indicated that the estimates are both accurate and precise. In addition, the proportion of users with a positive satisfaction and its variance approached the true value and zero, respectively, which indicated that the estimator is also consistent.

Overall, it was shown that social media data can be utilized to come up with estimates that can be instrumental in the improvement of the services of an agency.

Literature Cited

BAIG, A., DUA, A., & RIEFBERG, V. (2014). PUTTING CITIZENS FIRST: How to improve citizens' experience and satisfaction with government services. McKinsey Center for Government. Retrieved September 15, 2018, from https://www.mckinsey.com/~/media/mckinsey/industries/public_sector/our_insights/how_us_state_governments_can_improve_customer_service/putting_citizens_first_how_to_improve_citizens_experience_and_satisfaction_with_government_services.ashx

CHEN, Y., & SKIENA, S. (2014). Building Sentiment Lexicons for All Major Languages. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 383-389. doi:10.3115/v1/p14-2063

FELLOWS, I. (2018). wordcloud: Word Clouds. R package version 2.6. Retrieved April 8, 2019, from https://cran.r-project.org/web/packages/wordcloud/index.html

HU, M. & LIU, B. (2004). Mining and summarizing customer reviews. KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 168-177. doi:10.1145/1014052.1014073.

KEARNEY, M. W. (2018). rtweet: Collecting Twitter Data. R package version 0.6.8. Retrieved October 10, 2018, from https://cran.r-project.org/web/packages/rtweet/index.html

MULLINS, K. (n.d.) How to Improve Customer Satisfaction with Government Services. BrainStorm Group, Inc. Retrieved September 15, 2018, from http://www.dbizinstitute.org/resources/articles/how-improve-customer-satisfaction-government-services

PORCALLA, D. (2018, June 11). Top internet users: Philippines ranks 12th. Philippine Star. Retrieved September 1, 2018, from https://www.philstar.com/headlines/2018/06/11/1823578/top-internet-users-philippines-ranks-12th