# 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

Institute of Statistics
University of the Philippines Los Baños



#### **Consumer Satisfaction**

Collecting feedback to meet consumers' expectations



Real time data

Alternative to traditional methods?

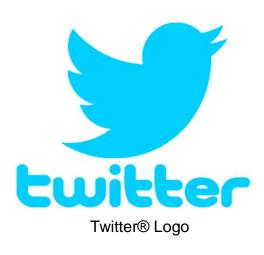




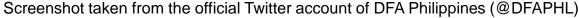


#### **Objective**

#### develop an alternative methodology to determine the satisfaction of users









# **Significance**







Data availability

# **Scope and Limitations**



November 4, 2018 to February 4, 2019



Casual nature of messages



Bias on specific age groups



Actual feelings of a person



Dictionary-based



#### Methodology

Gathering tweets using the Twitter API

Specifying the username of the account to be investigated

Determining overall sentiment score of a tweet

- Updating of the dictionary (optional)
- Matching the words with an existing dictionary

Filtering and preprocessing of data

- Excluding irrelevant or spam tweets
- Excluding tweets not written in English nor Tagalog
- Retaining only one tweet per use

Categorizing tweets based on sentiment

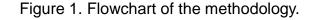
 Categorizing into positive, neutral, or negative based on sentiment score

Cleaning and tokenizing of tweets

Removing elements of a tweet that are not words

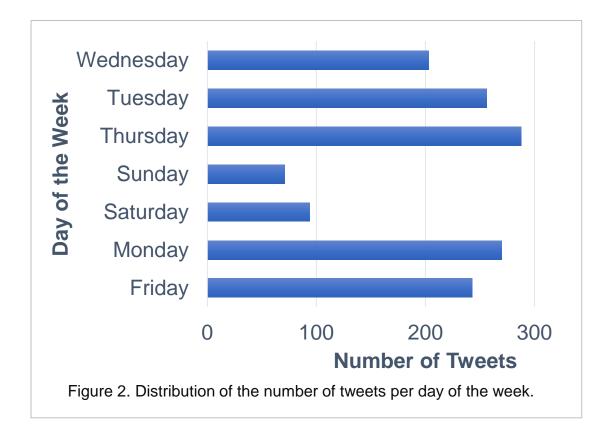
Computing the proportion of positive tweets

 Investigating the contents of the tweets tagged as negative





#### **Tweet Characteristics**





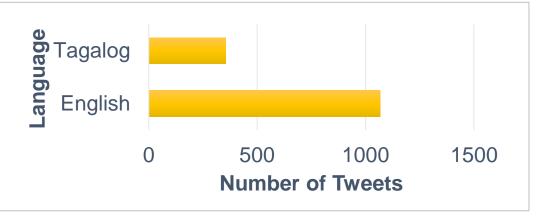
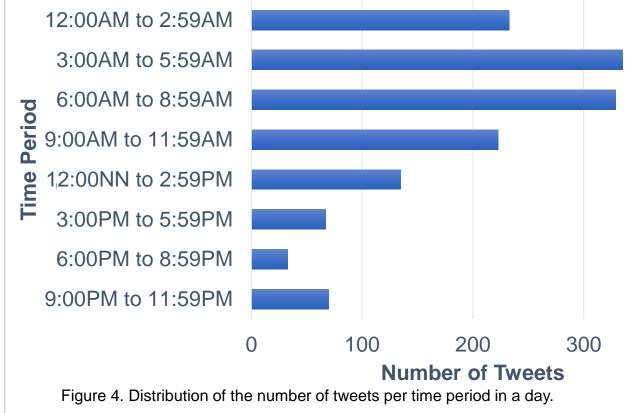


Figure 3. Distribution of the number of tweets per language.



#### **Word Clouds** number embassy certification payment renewal last lane online anak dont get old time psa newhelp visa scheduled aseana please certificate reference schedule months ibang tapos thanks application kukuha nso namebirth process office application now renew date online philippines requirements process Tagalog English

Figure 5. Word cloud of the tweets by language.



#### **Polarity of Words per Tweet**

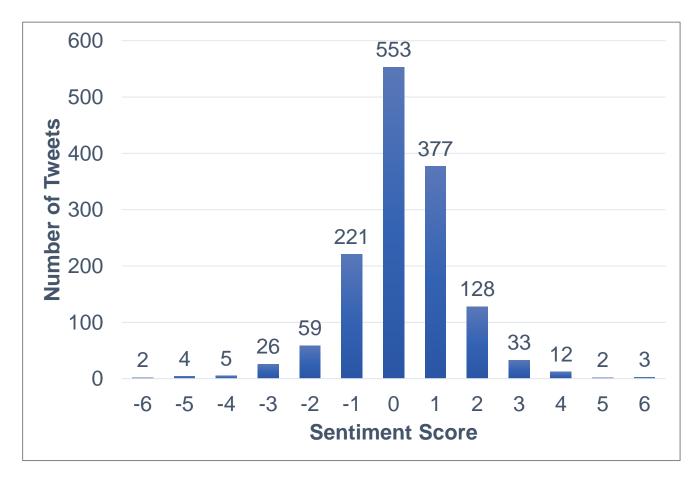


Figure 6. Distribution of the sentiment scores of the tweets.



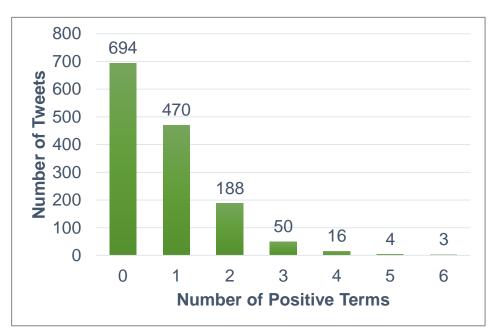


Figure 7. Distribution of the number of positive terms per tweet.

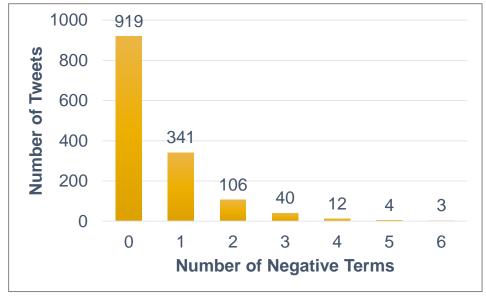


Figure 8. Distribution of the number of negative terms per tweet.

#### **Associations with the Sentiment Score**

Table 1. Measures of association of the sentiment scores with the characteristics of the tweets.

CHARACTERISTIC OF A TWEET	MEASURE USED	VALUE OF COEFFICIENT
Number of Words in a Tweet	Pearson's Correlation Coefficient	-0.1418
Language (English=0 ; Tagalog=1)	Point Biserial	-0.1258
AM/PM (AM=0; PM=1)	Point Biserial	-0.0522
Day of the Week	Eta Coefficient	0.0021



# **Sentiment Category**

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

SENTIMENT CATEGORY	FREQ	%
Positive	304	38.92
Neutral	297	38.03
Negative	180	23.05
Total	781	100.00

# **Bootstrapping**

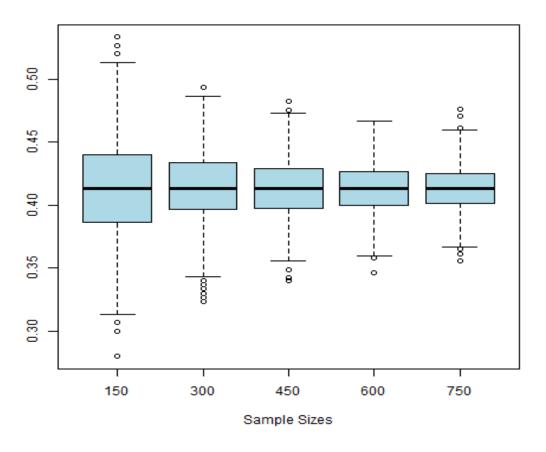


Figure 9. Distribution of the proportion of users with a positive satisfaction at different sample sizes with bootstrap resample of B=1000.



#### **Summary of the Research**

- ✓ Use of publicly posted Twitter data
- ✓ Tweet characteristics
- ✓ Sentiment scores of tweets
- ✓ Statistical properties of the estimates

#### **Recommendations for Future Researches**

- ✓ Other agencies
- ✓ A more comprehensive sentiment analysis
  - ✓ dictionary and word groupings
- ✓ Longer time period
- ✓ Wider scope



# Thank you!

