IFTTT

Sentiment Analysis of Asian American Tweets During the COVID-19 Pandemic and the Surge of Asian-American Hate Crimes

Jonathon Sun |GSE: Higher Education

Ying Dai | School of Nursing: Nursing Science

Jiayi Zhang | GSE: Learning Sciences & Technologies

Introduction

Asian American hate crimes has been on the rise since the beginning of covid-19 with rhetoric such as Kung-flu and the China virus. Due to the current social climate, twitter data using the search term Asian American is more crucial than ever to understand how Asian Americans are framed within the broader racial context of the United States.
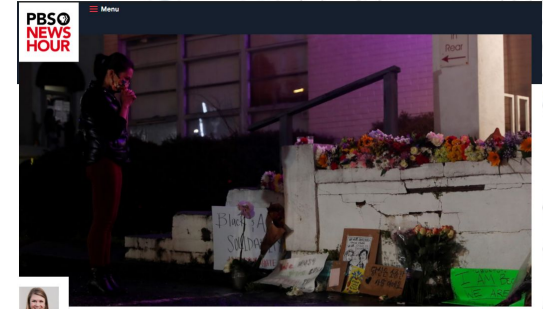

What we know about the Atlanta spa shootings that killed 8, including 6 Asian women

Opinion: I'd be reveling in Asians' success at the Oscars if I weren't so concerned for our safety

Trump Defends Using 'Chinese Virus' Label, Ignoring Growing Criticism

"It's not racist at all," the president said. But experts warned that the term could result in xenophobia.

Covid 'hate crimes' against Asian Americans on rise

Asian American businesses are defending themselves against rise in anti-Asian violence
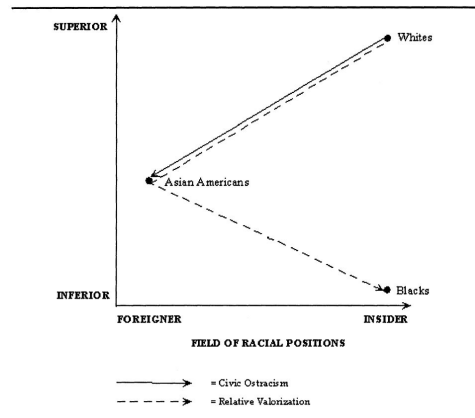
There's an economic cost to racism as business owners reduce hours, shell out for security in the wake of the Atlanta shootings

Background: Asian American Racial Positionality in the US

The Model Minority

Problematic in that…

(1) Frames Asian Americans as universally successful and reinforce color blind ideologies

(2) assumes Asian Americans are problem free and do not require resources and support

(3) pits Asian Americans against other minority groups.

(Yi and Musues, 2015).

Racial Triangulation (Kim, 1999)

Tri-Racial System (Bonilla-Silva, 2004)

Research Question

How is the issue of the model minority relevant and pertinent to a data science method?

Using data from Asian-American related tweets over the past five-months Twitter data between November 30th, 2020 2020 to April 1st, 2021 were collected using a tool called " If This, Then That (IFTTT)" (**https://ifttt.com/**)**.** We propose to use text mining, lasso, logistic regression, and random forest to conduct our analysis.

1.   Explore how Asian Americans are framed and described in social media during the COVID-19 pandemic
2.   Predict what common words are most likely to increase the likelihood of a tweet being retweeted on the topics around Asian Americans.

Data Collection + Cleaning + Wrangling

Data Collection	
- Used IFTTT - Pulled twitter data that contains #AsianAmerican #AAPI - For the time period of 3/1/2020 - 3/1/2021	The datasets contains - 126,622 tweets in total - 32,135 unique tweets - 7 variables (text, time, date, username, etc.)

The grandmother who fought off an attacker in San Francico has decided to donate ALL #gofundme money back to the #AAPI community to fight racism.
Her @gofundme raised nearly $900,000. $900,000!!
We can all learn a lot from Xia Zhen Xie. Incredible.

@nbcbayarea
#StopAsianHate🔥

Data Cleaning	
Created new features -	- Number of retweets - Whether a tweet was retweeted (Binary)
Within each tweet, we removed -	- Hyperlinks - @ - # - Punctuations - Digits/numbers - 😊 - Whitespace - &amp
For each word, we assigned -	- Binary sentiment (positive/negative) - Sentiment type/ affect (10 sentiments; e.g. Anger, anticipation, disgust…) - Sentiment score

RQ 1 results

1. Explore how Asian Americans are framed and described in social media during the COVID-19 pandemic

- Emotion classification using NRC Word-Emotion Association Lexicon (aka EmoLex)

- Eight emotions
  - Highest emotions: fear, followed by trust, anger, and sadness...

- Overall positive and negative emotions:
  - 56% positive emotions

- Syuzhet package
  - get_nrc_sentiment()

RQ 1 results

1.   Explore how Asian Americans are framed and described in social media during the COVID-19 pandemic
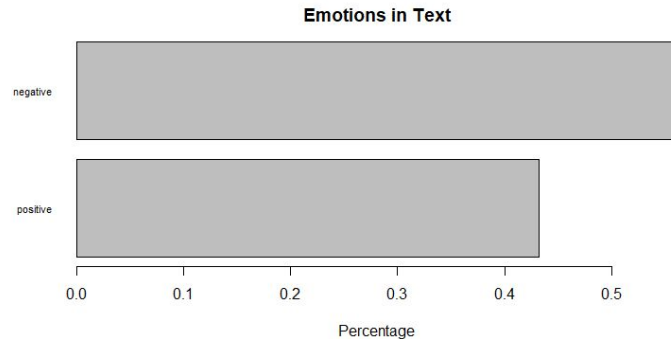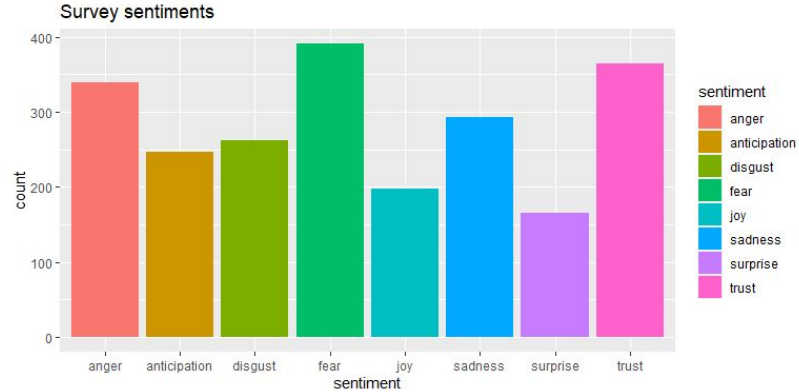
RQ 2 results

2. Predict what common words are most likely to increase the likelihood of a tweet being retweeted on the topics around Asian Americans.

Analysis 1: LASSO, Logistic regression, Word clouds and Predictive Power
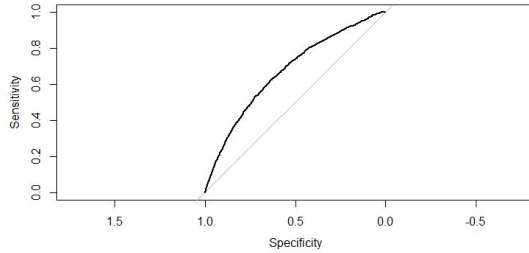
Misclassification Error	0.18
Area Under Curve	0.672

Analysis 2: Neuronet

◎ **Two layers with 16 and 8 neurons in each layer**

Loss	0.510
Accuracy	0.751

RQ 2 results: Word Clouds

In total there were 86 words that were indicators of if a tweet was more likely to be retweeted or not.

Retweet

Good words: 67

No Retweet

Bad words: 18

Conclusion

RQ1: Explore how Asian Americans are framed and described in social media during the COVID-19 pandemic

Many of the texts that we analyzed showed a sense of fear and anger; however, there were also many tweets about trust.

RQ2: Predict what common words are most likely to increase the likelihood of a tweet being retweeted on the topics around Asian Americans.

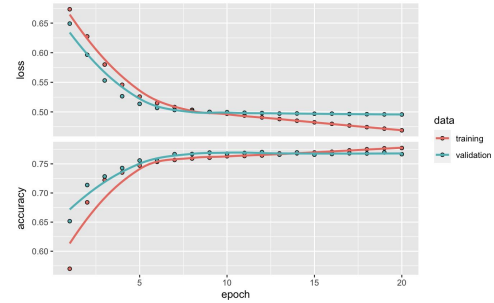When words about tweets were analyzed, retweets that were tweeted out were about leading and engaging in discussion. Tweets that were not likely to be retweeted were about specific racial and ethnic groups.

Retweet

No Retweet