Examining Attitudes Towards Face Masks Using Sentiment Analysis of Tweets

Shriya Iyer
under the mentorship of Dr. Lauren Cappiello
Assistant Professor of Mathematics and Statistics
California State University

August 2020

Abstract

Currently, the world is battling against the COVID-19 pandemic. First detected in Wuhan, China in December 2019, this infectious disease has affected more than 21 million people worldwide [3]. Because COVID-19 is a respiratory illness, epidemiologists and public health experts recommend that individuals wear face masks to contain small respiratory droplets that come out of one's mouth and/or nose when talking, sneezing, or coughing [13]. Data supports wearing face masks as a measure to slow the spread of COVID-19 [9]. Unfortunately, disinformation being spread about face masks, including that they cause carbon dioxide poisoning and low oxygen levels, threatens their usage [10]. This research project aims to examine personal attitudes towards face masks using social media posts extracted from Twitter. Using Python and the Twitter API to extract one hundred million Tweets, it was found that on average, 60% of Tweets related to face masks were neutral. Therefore, this study concludes that the average public sentiment towards face masks is neutral.

1 Introduction

1.1 Face Mask Usage

COVID-19 is a mild to severe respiratory illness caused by a novel coronavirus [1]. The virus that causes this infectious disease is spread from person to person through respiratory droplets produced when an infected person coughs, sneezes, or talks [2]. Face masks serve as a barrier by preventing these droplets from traveling into the air and on to other people [4]. They are a critical preventative measure that reduce the spread of illness.

Figure 1: A graphic showing the various types of face masks. [6]

1.2 Misinformation

Unfortunately, misinformation threatens public beliefs in the usefulness of face masks to slow the spread of COVID-19. One common misconception is that wearing face masks deprives your body of oxygen, a condition known as hypoxia. The World Health Organization (WHO) has addressed this misconception and stated that the prolonged use of medical masks when properly worn does not cause oxygen deficiency [10]. Other misbeliefs include that face masks cause carbon dioxide poisoning and harm the immune system.

Misinformation about face masks has caused polarizing opinions about their effectiveness. Understanding whether the average attitude towards face masks is neutral, negative, or positive is important for two reasons: 1) it could indicate how pervasive and accepted disinformation is and 2) it could serve to explain the continued spread of COVID-19.

Figure 2: A viral misleading graphic about the dangers of long-term face mask usage. [7]

2 Literature Review

Running sentiment analysis on Tweets is a common method used by researchers to determine personal attitudes towards a subject, especially in regards to a disease or illness. This is because social media offers a platform for individuals to not only share information about the disease or illness but also share their opinions and experiences with it.

Numerous studies that utilize Twitter analysis have been conducted in relation to COVID-19. One study done by Cornell University aimed to analyze misinformation in Twitter conversations. They utilized the Twitter application programming interface (API) to collect data and tracked emerging sentiments and hashtags over countries [15]. They also tracked sentiments regarding social distancing and work from home hashtags. Another study was able to quantify these sentiments by examining Tweets related to COVID-19 from the WHO and the general public since January 2020. Their results showed that the majority of Tweets had a positive polarity, with only 15% of them being negative [14].

Apart from generalized sentiments towards COVID-19, several studies have also explored attitudes towards specific public impacts that the disease has caused. A study done by the American Geriatrics Society analyzed Tweets related to COVID-19 and older adults in order to identify ageist content that implied the life of older adults was less valuable. They obtained a representative sample of original Tweets using multiple keywords and hashtags, including "elderly", "boomer", and #coronavirus. Their results showed that almost one-quarter of analyzed Tweets had ageist or potentially offensive content towards older adults [11]. Additionally, another study that ran sentiment analysis on Tweets related to reopening in the US found that people had a less negative sentiment towards the situation of reopening in comparison to the lockdown situation [8].

Although all of the mentioned studies examined Tweets and employed sentiment analysis, none of them looked specifically at attitudes towards face masks. Indeed, there are not many studies about this topic in current literature. Therefore, this research project aims to solve for this information and knowledge gap.

3 Methods

3.1 Extracting Tweets

Python programming was used to extract data from Twitter. The Python library Tweepy was set to access the Twitter API and gather 10,000,000 Tweets. The count of 10,000,000 Tweets was chosen since this provided a significantly large sample size. Using the query search on Tweepy, only Tweets that had the term ”face mask” or ”face masks” were extracted to ensure relevance to the research topic. Each Tweet that was gathered was cleaned by removing links and special characters. Then, the remaining text of the Tweet was appended to a set titled tweets. To ensure that retweets were not included, code was written so that each Tweet was only appended to the set once.

3.2 Sentiment Classifier

A Python sentiment classifier found on GeeksforGeeks was used to analyze the sentiment of each Tweet [5]. The sentiment classifier classifies Tweets by the following three attitudes: positive, negative, or neutral. It does this by assigning each Tweet a polarity between -1.0 and 1.0. Tweets that are greater than zero are labeled as positive, and those that are less than zero are labeled as negative. A Tweet with zero polarity is classified as neutral. The sentiment classifier is able to distinguish between the three attitudes by using training data from a Movie Reviews dataset. In the dataset, each review has already been labeled as positive or negative. The positive and negative features extracted from the movie reviews are used by the sentiment classifier to determine the positive and negative attributes of each Tweet.

3.3 Analyzing Sentiments

A percentage of positive, negative, and neutral tweets was computed by dividing the number of positive, negative, and neutral Tweets by ten million, the total number of Tweets extracted. Since Tweepy extracts the most recent Tweets, the code was run ten times to ensure accuracy and expand the variety of Tweets included. The sentiment percentages calculated per run were averaged.

4 Results

The results of the ten runs show a fairly consistent trend among the three sentiment groups. The percentage of positive Tweets was centered around approximately 30%, the percentage of neutral Tweets was centered around 60%, and the percentage of negative Tweets was centered around approximately 10%. Neutral tweets were the most prominent followed by positive tweets.

	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Run 10	Average
% Positive Tweets	31%	28%	28%	30%	32%	29%	28%	29%	30%	33%	29.8%
% Neutral Tweets	59%	61%	62%	59%	58%	60%	60%	61%	61%	59%	60.0%
% Negative Tweets	10%	11%	10%	11%	10%	11%	12%	10%	9%	8%	10.2%

Figure 3: The table displays the sentiment percentages for each of the ten runs and the final average sentiment percentages.

5 Discussion

5.1 Sentiment Analysis

The results showed that the average public sentiment towards face masks is neutral. Although there was no specific hypothesis as to what would be the prominent sentiment, it was expected that there would be a large percentage of negative Tweets due to misinformation. However, the results showed that only 10% of Tweets on average had negative sentiments towards face masks, which was much lower than expected.

The large percentage of neutral Tweets could be explained by a number of factors. First, the immense amount of news shared on Twitter could have skewed the neutral sentiment percentage. It was observed that many of the Tweets classified as neutral were not necessarily personal opinions but simply the sharing of a news article.

Furthermore, the high percentage of neutral sentiments could also be explained by the failure of the sentiment classifier. For example, a Tweet that read,

"The Fight for Your Liberty: Face Masks
Fight the Democrat/ Deep State Cult of the Mask"

was classified as neutral. However, the connotations associated with this Tweet and words like "liberty" clearly indicate a negative sentiment towards face masks. Another neutral Tweet read,

"My sister Kelly, who already has lupus,
has COVID-19. This is her hospital room.
1 of you Non-Mask wearers did this."

This Tweet clearly describes the hospitalization of an individual due to "Non-Mask wearers". Although this Tweet demonstrates a negative tone, it is evident that the author is upset with non-mask users. Therefore, they appear to support face mask usage. Tweets like this are difficult for the sentiment classifier to group into the positive, negative, or neutral attitude groups. Thus, the failure of the sentiment classifier to accurately analyze the sentiments of each Tweet most likely contributed to a large percentage of neutral Tweets.

5.2 Suggestions

It is suggested that if sentiments of Tweets relating to face masks are analyzed, then a machine learning model should be integrated into the sentiment classifier. This will allow the sentiment classifier to learn and improve when identifying positive and negative attributes of each Tweet. If a machine learning model is not used, then the sentiment classifier should have better training data to work with, specifically a dataset that has already labeled Tweets as positive or negative. This will allow the sentiment classifier to be more accustomed to jargon and terminology found on social media.

Furthermore, it is suggested that when analyzing sentiments, data should only be extracted from a particular geographic region. For example, sentiments in particular cities, states, or countries would provide more insightful information to public health experts and policymakers. Since this research project analyzed Tweets from any location, the sentiments shown are from a global perspective, making it harder to provide reasoning or prescribe any meaningful course of action.

6 Conclusion

With face masks serving an important role in slowing the spread of COVID-19, it is important that any negative sentiments towards them are addressed. The aim of this research project was to examine average sentiments towards face masks using Tweets extracted from Twitter. Using Python and the Twitter API to extract one hundred million Tweets, it was found that on average, 60% of Tweets related to face masks were neutral. Therefore, this study concludes that the average public sentiment towards face masks is neutral.

With the COVID-19 vaccine still in development, it is inevitable that the usage of face masks will continue for a long period of time [12]. However, the results of this research project indicate that there is no overwhelming opposition to wearing them. In fact, there are more positive attitudes from individuals in comparison to negative ones. Nevertheless, the presence of negative sentiments is still an issue that needs to be addressed.

Analyzing sentiments over time and matching them with major pandemic events is a point of interest for future research. Looking at how sentiments change over time can help to establish trends and predict how the public will react to certain events. For example, seeing if negative sentiments towards face masks increase as a country or state increases in the number of COVID-19 cases would be interesting to study. Ideally, the sentiment data will allow policymakers and public health experts to find ways to address the public more effectively and further encourage the usage of face masks.

References

[1] Definition of COVID-19 by Merriam-Webster. https://www.merriam-webster.com/dictionary/COVID-19. (Accessed on 08/09/2020).

[2] Coronavirus (COVID-19) Frequently Asked Questions. https://www.cdc.gov/coronavirus/2019-ncov/faq.html: :text=TheAugust 2020. (Accessed on 08/10/2020).

[3] Coronavirus Map: Tracking the Global Outbreak. https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html, August 2020. (Accessed on 08/13/2020).

[4] COVID-19: Considerations for Wearing Masks. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html, August 2020. (Accessed on 08/10/2020).

[5] Twitter Sentiment Analysis using Python. https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/, July 2020. (Accessed on 08/13/2020).

[6] Type of face masks concept vector. https://www.freepik.com/free-vector/type-face-masks-concept$_8845420.htm, June 2020. (Accessed on 08/12/2020)$.

[7] AFP. Misleading Graphics About Dangers Of Long-term Mask Use Viral. https://www.boomlive.in/world/misleading-graphics-about-dangers-of-long-term-mask-use-viral-8594, June 2020. (Accessed on 08/11/2020).

[8] Mohammed Emtiaz Ahmed, Md Rafiqul Islam Rabin, and Farah Naz Chowdhury. Covid-19: Social media sentiment analysis on reopening. *arXiv preprint arXiv:2006.00804*, 2020.

[9] Nina Bai. Still Confused About Masks? Here's the Science Behind How Face Masks Prevent Coronavirus. https://www.ucsf.edu/news/2020/06/417906/still-confused-about-masks-heres-science-behind-how-face-masks-prevent, July 2020. (Accessed on 08/09/2020).

[10] Jack Goodman and Flora Carmichael. Coronavirus: 'Deadly masks' claims debunked. https://www.bbc.com/news/53108405, July 2020. (Accessed on 08/10/2020).

[11] Maria Renee Jimenez-Sotomayor, Carolina Gomez-Moreno, and Enrique Soto-Perez-de Celis. Coronavirus, ageism, and twitter: An evaluation of tweets about older adults and covid-19. *Journal of the American Geriatrics Society*, 2020.

[12] Olga Khazan. How Long Will We Have to Wear Masks? - The Atlantic. https://www.theatlantic.com/health/archive/2020/05/pandemic-masks-face-shields/611971/, May 2020. (Accessed on 08/15/2020).

[13] Lisa Lockerd Maragakis. Coronavirus Face Masks: Types & When to Use. https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/coronavirus-face-masks-what-you-need-to-know, August 2020. (Accessed on 08/09/2020).

[14] Nikhil Kumar Rajput, Bhavya Ahuja Grover, and Vipin Kumar Rathi. Word frequency and sentiment analysis of twitter messages during coronavirus pandemic. *arXiv preprint arXiv:2004.03925*, 2020.

[15] Karishma Sharma, Sungyong Seo, Chuizheng Meng, Sirisha Rambhatla, Aastha Dua, and Yan Liu. Coronavirus on social media: Analyzing misinformation in twitter conversations. *arXiv preprint arXiv:2003.12309*, 2020.

Appendix

Python Code

```
import re
import tweepy
from tweepy import OAuthHandler
from textblob import TextBlob

class TwitterClient(object):
    '''
    Generic Twitter Class for sentiment analysis.
    '''
    def __init__(self):
        '''
        Class constructor or initialization method.
        '''
        consumer_key =''
        consumer_secret =''
        access_token =''
        access_token_secret =''

        # attempt authentication
        try:
            # create OAuthHandler object
            self.auth = OAuthHandler(consumer_key, consumer_secret)
            # set access token and secret
            self.auth.set_access_token(access_token, access_token_secret)
            # create tweepy API object to fetch tweets
            self.api = tweepy.API(self.auth)
        except:
            print("Error: Authentication Failed")

    def clean_tweet(self, tweet):
        '''
        Utility function to clean tweet text by removing links, special
        characters using simple regex statements.
        '''
        return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)", " ", tw

    def get_tweet_sentiment(self, tweet):
        '''
        Utility function to classify sentiment of passed tweet
        using textblob's sentiment method
        '''
        # create TextBlob object of passed tweet text
```

```
        analysis = TextBlob(self.clean_tweet(tweet))
        # set sentiment
        if analysis.sentiment.polarity > 0:
            return 'positive'
        elif analysis.sentiment.polarity == 0:
            return 'neutral'
        else:
            return 'negative'

    def get_tweets(self, query, count = 10):
        '''
        Main function to fetch tweets and parse them.
        '''
        # empty list to store parsed tweets
        tweets = []

        try:
            # call twitter api to fetch tweets
            fetched_tweets = self.api.search(q = query, count = count)

            # parsing tweets one by one
            for tweet in fetched_tweets:
                # empty dictionary to store required params of a tweet
                parsed_tweet = {}

                # saving text of tweet
                parsed_tweet['text'] = tweet.text
                # saving sentiment of tweet
                parsed_tweet['sentiment'] = self.get_tweet_sentiment(tweet.text)
                parsed_tweet['location'] = tweet.user.location
                parsed_tweet['createdate'] = tweet.created_at
                print(tweet.user.location)

                # appending parsed tweet to tweets list
                if tweet.retweet_count > 0:
                    # if tweet has retweets, ensure that it is appended only once
                    if parsed_tweet not in tweets:
                        tweets.append(parsed_tweet)
                else:
                    tweets.append(parsed_tweet)

            # return parsed tweets
            return tweets

        except tweepy.TweepError as e:
            # print error (if any)
```

```python
        print("Error : " + str(e))

def main():
    # creating object of TwitterClient Class
    api = TwitterClient()
    # calling function to get tweets
    tweets = api.get_tweets(query = 'face masks OR face mask', count = 1000000)


    # picking positive tweets from tweets
    ptweets = [tweet for tweet in tweets if tweet['sentiment'] == 'positive']

    # percentage of positive tweets
    print("Positive tweets percentage: {} %".format(100*len(ptweets)
    /len(tweets)))

    # picking negative tweets from tweets
    ntweets = [tweet for tweet in tweets if tweet['sentiment'] == 'negative']

    # percentage of negative tweets
    print("Negative tweets percentage: {} %".format(100*len(ntweets)
    /len(tweets)))

    # percentage of neutral tweets
    print("Neutral tweets percentage: {} %".format(100*(len(tweets)
    -(len( ntweets )+len( ptweets)))/len(tweets)))
    nutweets = [tweet for tweet in tweets if (tweet['sentiment'] != 'negative'
    and tweet['sentiment'] != 'positive')]

    # printing first 5 positive tweets
    print("\n\nPositive tweets:")
    for tweet in ptweets[:10]:
        print('tweet text:',tweet['text'], 'tweet location:',tweet['location'],
        'tweet date:', tweet['createdate'])

    # printing first 5 negative tweets
    print("\n\nNegative tweets:")
    for tweet in ntweets[:10]:
        print(tweet['text'])


    # printing first 5 neutral tweets
    print("\n\nNeutral tweets:")
    for tweet in nutweets[:10]:
        print(tweet['text'])
```

```python
if __name__ == "__main__":
    # calling main function
    main()
```