I2.01 Evaluating the Effect of Misinformation Within Industries

Jesse Contreras, Brandon McCarthy, Robert Torres
Ingram School of Engineering, Texas State University

Project Goal

- To understand populations' interaction with online misinformation and how it affects their workflow. By seeing how misinformation impacts people of different industries, methods can be developed to combat and mitigate incorrect information online.
- To label input data for a machine learning algorithm based on how factual they are.
- Create and conduct interviews within the industries of medical, small businesses and higher education.

Background Information

" I think history will remember that it wasn't just the blatant misinformed narratives that were put out by some scientist that caused huge damage to the pandemic – but also the face that media outlets continued to platform them well after they had been repeatedly wrong. " - Anonymous Twitter User

Figure 1. Example Tweet

- As of April 2022, worldwide there have been a total of 510 million reported COVID-19 cases. The United States is leading the count with 81 million cases and 990 thousand recorded deaths.
- It has been shown that misinformation online can and has led to vaccine hesitancy, people believing in false remedies, and people not taking the proper precautions when in contact with the contagious virus.
- According to Pew Research Center, 54% of Americans get their news from social media sources.

Project Purpose

- The purpose of this project is to first gather data in order to identify the amount of misinformation on COVID-19 within the social media platform twitter.
- Interviews are conducted to measure the amount of misinformation identified by professionals within the medical field, small businesses and higher education.
- Through these objectives we will then determine whether further research needs to be conducted in hopes to lessen the effects and reach of misinformation and the grasp it has on our society.

Labeling

- 15,000 tweets labeled as input data for a machine learning algorithm.

Machine Learning Algorithm Labels:

- -1 false information
- -0.5 partially false
- 0 neutral / irrelevant
- 0.5 partially true
- 1 true information

- Tweets were crosschecked once by another team member. If the labelers agreed the label stays.
- An accuracy score will be outputted from the machine learning algorithm to track performance.
- A VBA macro was used to help speed up the labeling process. Increased labeling speed by 30%.

Tweet Label Distribution

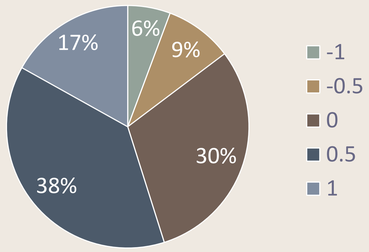


Figure 2. Tweet Labeling Resulting Distribution

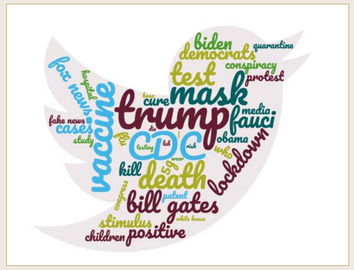
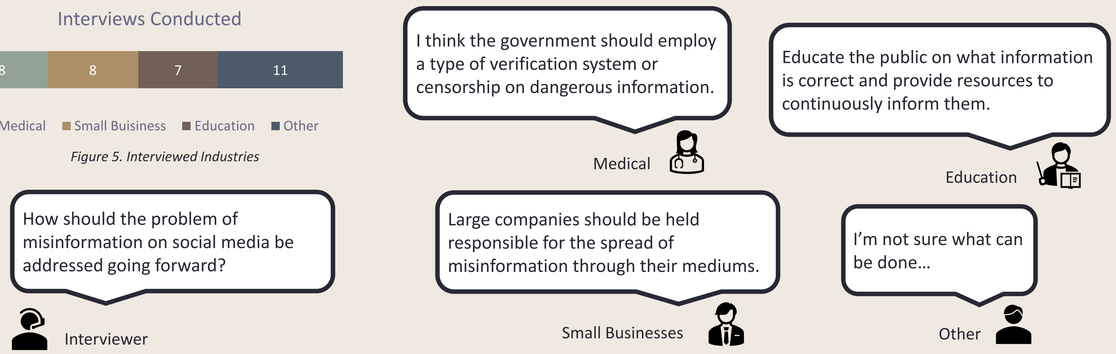


Figure 3. Labeling Dataset Most Occurring Keywords

Figure 4. Labeling VBA Macro GUI Created

Interviews

Figure 5. Interviewed Industries



Analysis

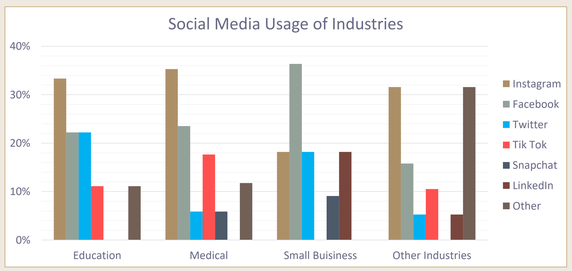


Figure 6. Social Media Usage Between Industries Chart

Table 1. Social Media Usage Between Platforms Table

Social Media	Usage
Instagram	59%
Facebook	44%
Twitter	24%
Tik Tok	21%
Snapchat	6%
LinkedIn	9%
Other	29%

- Aggregate industry perceived interaction with misinformation
  - Scale 1-10
  - 1 being no impact
  - 10 being a large impact

Table 2. Perceived Impact of Misinformation

Industry	Career	Online
Education	5.75	4.00
Medical	6.38	5.13
Small Buisiness	6.14	7.14
Other	6.20	6.32
Average	**6.17**	**5.88**

Methodology

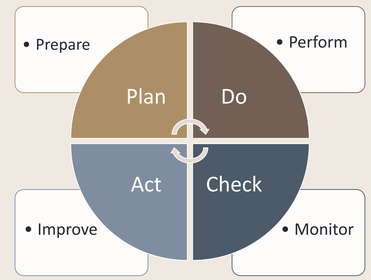


Figure 7. Plan Do Check Act Methodology

Human Factors

- External:
  - Can lead to inaccurate dispositions, misguided actions, and hindered operational ability.
- Internal:
  - Labeling a large number of tweets can lead to physical and mental fatigue.

Future Work

- Collect a larger sample of interviews
- Interview more people outside of Texas
- Interview more than three industries
- Work on methods to mitigate misinformation

The Project Team

Brandon McCarthy (Left), Jesse Contreras (Middle), Robert Torres (Right)

Acknowledgments

- We would like to thank the following individuals for their help:
  - Dr. Zhijie Dong, Texas State University
  - Dr. Londa, Texas State University
  - Alexander Little, Texas State University