

I 1.01- Evaluating the Spread of Misinformation on Social Media

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

Project Goal

- To identify people's interaction with misinformation on online regarding the COVID-19 pandemic. This way methods can be developed to combat and mitigate incorrect information online. Twitter will be the main social media platform evaluated.
- To label input data for a machine learning algorithm.
- To create and distribute a survey about where people receive their news to random sample of Texas residences.

Background Information

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

Figure 1. Example Tweet

- As of November 2021, there have been a total of 259 million recorded COVID-19 cases. The United States is leading the count with 4.31 million recorded cases.
- According to [Pew Research Center](#), 54% of Americans get their news from social media sources.
- It has been shown that misinformation online has and can lead to vaccine hesitancy, people believing in false remedies, and people not taking the proper precautions when in contact with the contagious virus.

Methodology

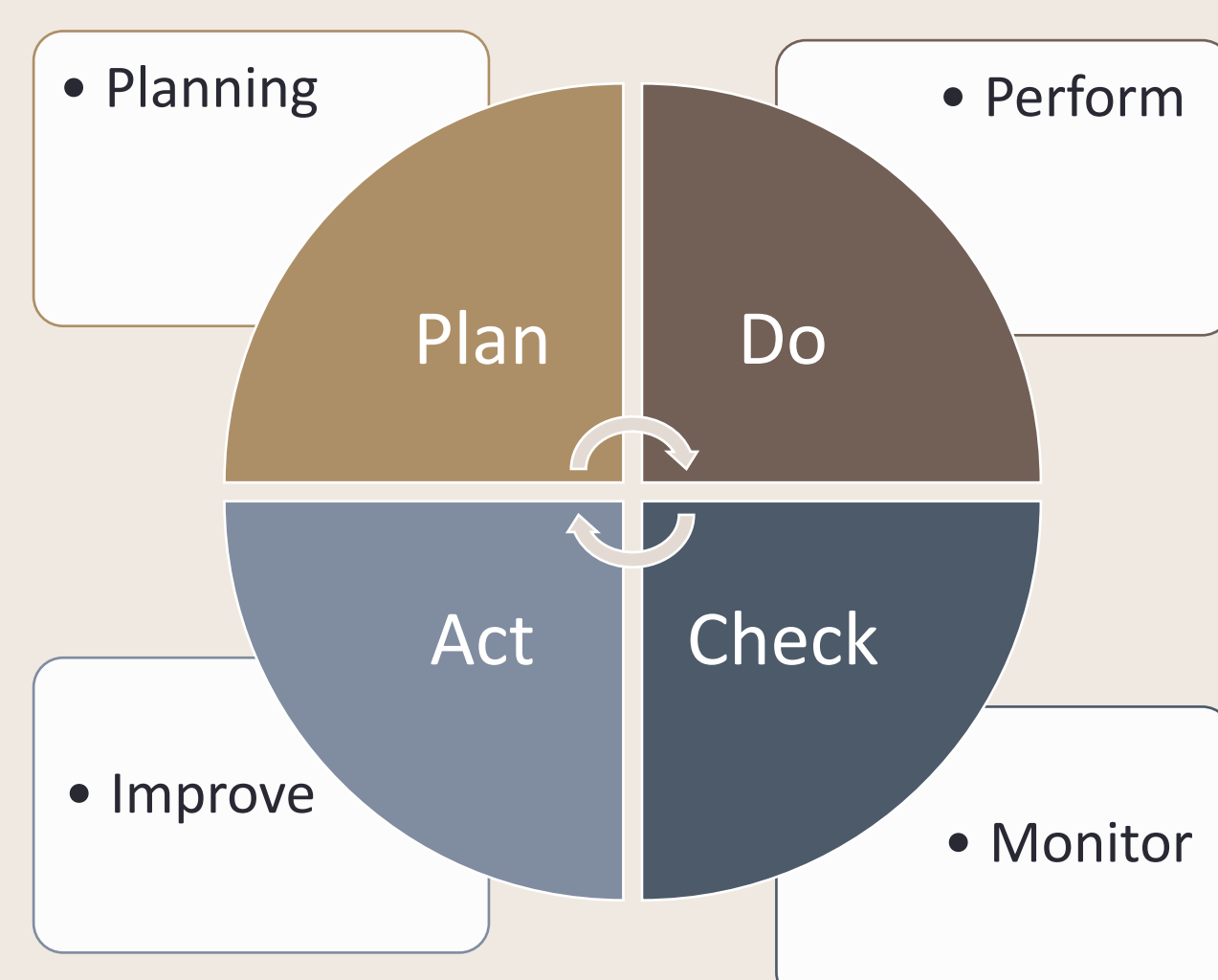


Figure 2. Plan Do Check Act Methodology

Labeling

- 15,000 tweets need to be labeled as input data for a machine learning algorithm.
- After all tweets are labeled, they must be crosschecked once with another team member. If the two labelers agree 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

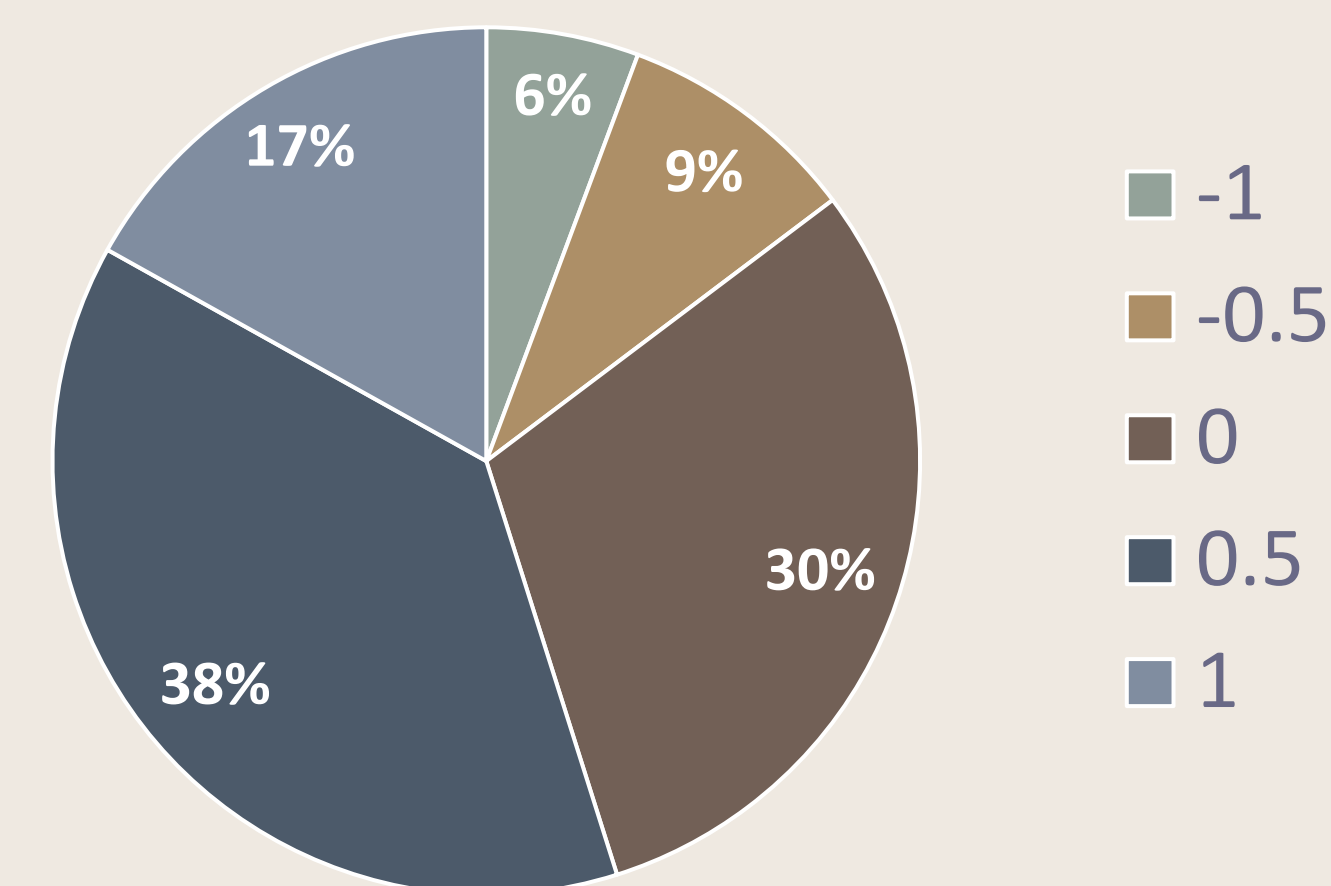


Figure 3. Tweet Labeling Resulting Distribution

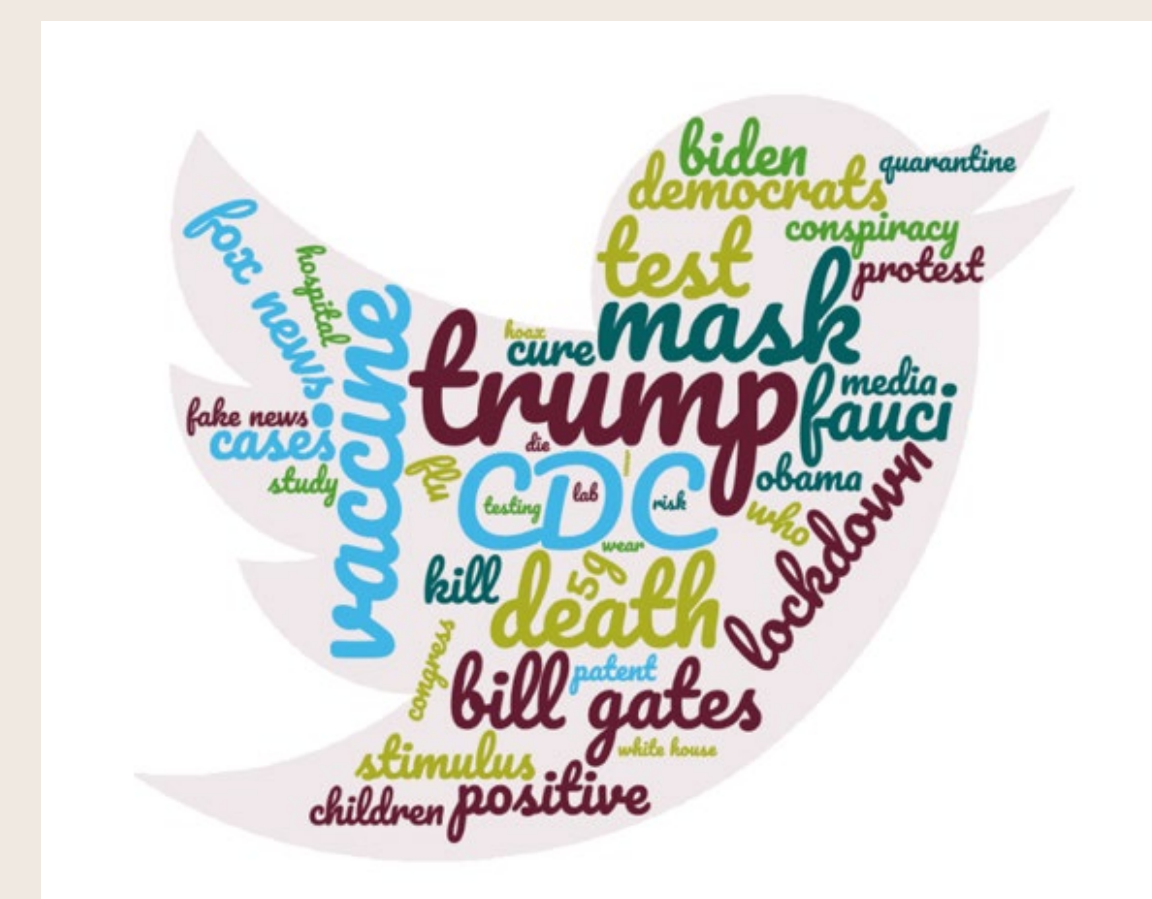


Figure 4. Labeling Dataset Most Occurring Keywords

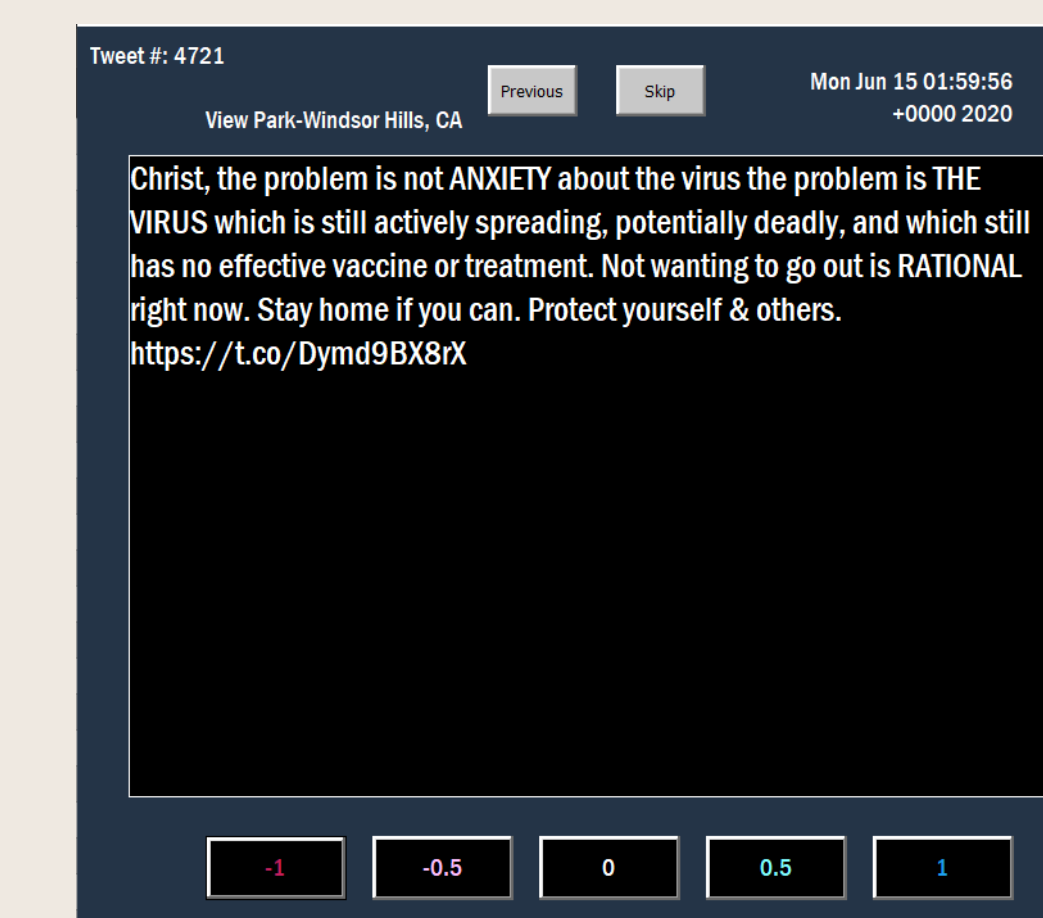


Figure 5. Labeling VBA Macro Created

Surveying

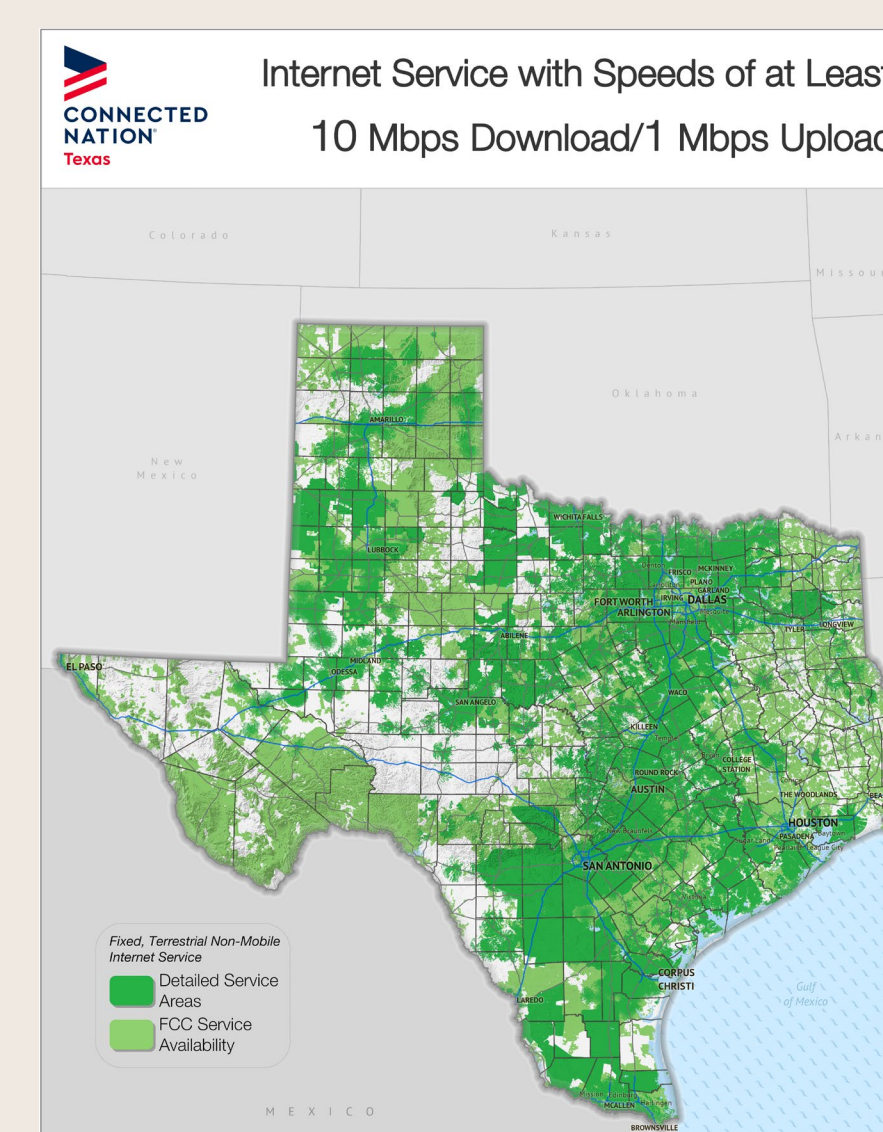


Figure 6. Internet Coverage Map in Texas

First Planned Surveying Regions:

- Texas State University
- Dallas
- Austin
- Houston

Survey Design:

- Distributed electronically
- Short and concise
- Unambiguous questions
- Disclosures of purposes and anonymity of the data collected

Metrics

Metric: Twitter Tweet Labeling	
Scale	Score
Tweets Labeled	
0 - 5,000	1
5,001 - 10,000	2
10,001 - 14,999	3
15,000	4

Table 1. Evaluation criteria for tweet labeling task

Metric: Survey	
Scale	Score
Survey Responses	
0 - 100	1
101 - 200	2
201 - 300	3
300+	4

Table 2. Evaluation criteria for surveying task

Metric: Report	
Scale	Score
Report Grade	
0 - 65%	1
66 - 80%	2
81 - 99%	3
100%	4

Table 3. Evaluation for research report task

Future Work

- Finish labeling the sets of tweets
- Crosscheck tweets
- Check accuracy of machine learning model
- Design Survey
- Distribute Survey
- Analyze Results
- Write Report

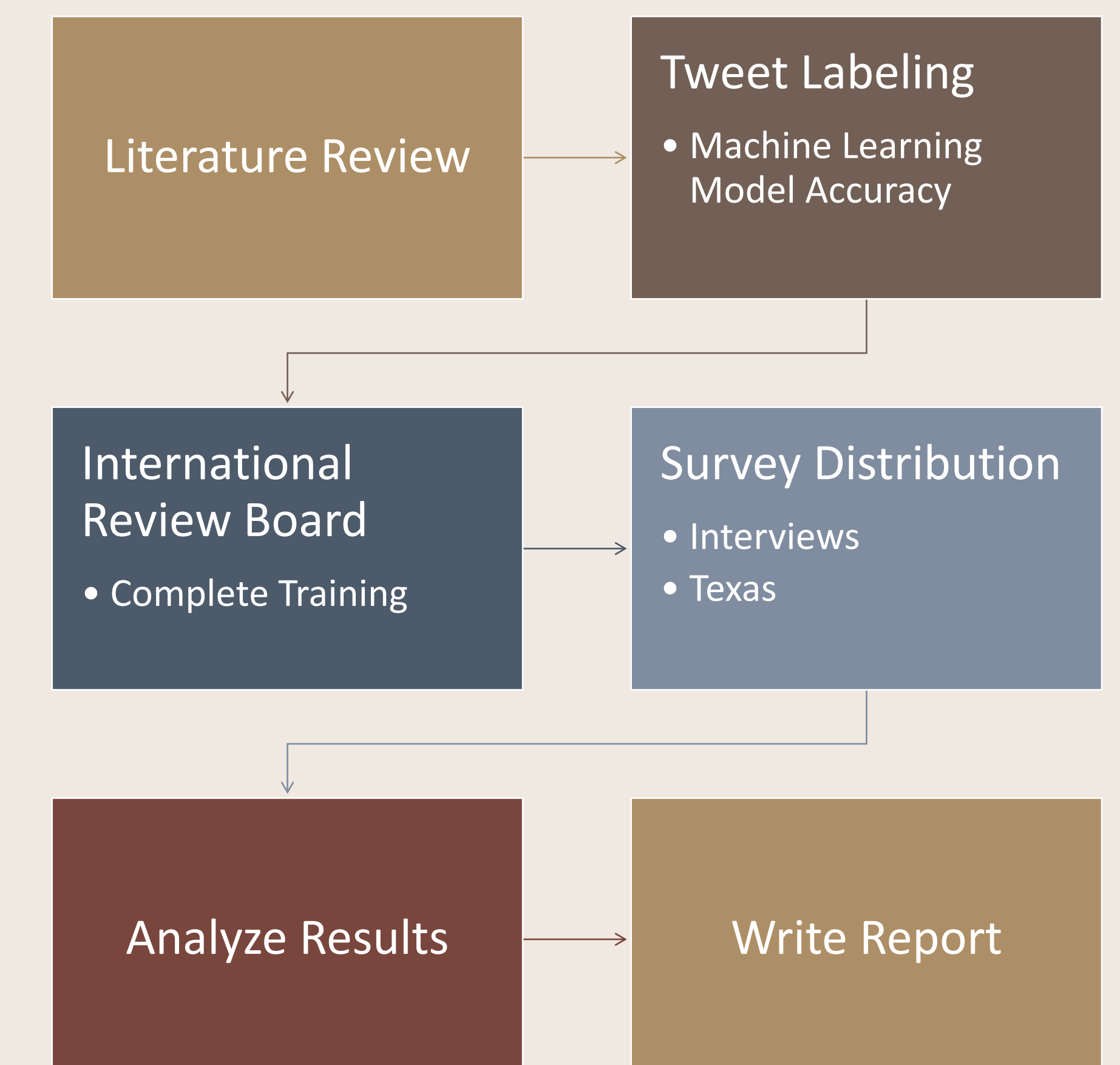


Figure 7. Project Progression Process

The Project Team



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