## Members: Shaoqian Chen, Jiahe Zhang, Tianxiao Yu, Zhounan(Madison) Wang

Motivating Background

Social Media has become an important part of our lives for many years. People nowadays are increasingly inclined to use social media platforms to express their ideas. Twitter, is considered one of the most successful social media platforms of our time. With over 145 million daily active users, Twitter provides us a vast amount of data everyday. In this project, we want to use data collected on Twitter to predict the outcome of the most trending topic at this time - the U.S presidential election scheduled for November 3rd, 2020.

Problem Statement

Given sets of positive, neutral and negative replies collected under presidential candidates’ most recent tweets, this project analyzes the sentiment of the public and makes a prediction about the outcome of the latest U.S. presidential election focusing on two candidates, Donald J. Trump and Joe Biden. Examples of what we mean by positive, neutral and negative replies can be found in Exhibit 1.

For whom are you solving this problem?

We are solving this problem for U.S. citizens and general public who pay attention to the presidential election analysts and would like to make sense of the sentiments of people towards them. It is like an online opinion poll. Ultimately, our result will be compared with the actual outcome to check the performance and feasibility of using social media to predict public opinions and thus replace traditional polling.

Why is this problem challenging?

First of all, we would like to point out that only textual data will be used in this project. Gathering other formats of data makes the task even more challenging. Besides, twitter provides freedom of speech to every user by giving them the opportunity to express themselves in various ways they desire. Thus, individuality in terms of speech and language might cause confusion, such as slang use, idioms and abbreviations, etc. Furthermore, one of the most important challenges is to find a way to deal with misleading information. Users may misspell a word or instead of expressing direct opinion they chose to be ironic about the topic.

## Datasets and Evaluation

In this part, we are planning to scrape replies of tweets done by both the runners using Twitter’s API. Ideally, we are considering scrape 5000 most relevant comments for both runners. Our main target will be the replies and time. Other data like username and @ wouldn’t add any value to our sentiment analysis.

For evaluation, we will focus more on using visualizing comparison with polls to show the accuracy or relevance. The seaborn’s distribution plot will be widely used for both candidates which is suitable for comparing range and distribution for groups of numerical data.

We are also considering assigning each comment a score and use metric like MAP to evaluate the result. More methods might come up after we get the full dataset.

Exhibits

| Sample Reply | Positive | Neutral | Negative |
| --- | --- | --- | --- |
| Donald Trump | This is what leadership looks like.  Thank you, President Trump. | How would we know? That just sounds like regular Trump | Hey CRAZY, you just put the BRAKES on the "DEAL" with a set of Tweets like an HOUR AGO  Have You TOTALLY LOST IT???? |
| The American people realize President Trump is doing everything possible to get things moving! | We still have time to register new voters. | TRUMP ELECTION LIE ALERT  Retweet to get [@Twitter](https://twitter.com/Twitter) to fact check this BS. |
| Joe Bidden | YEP!! Always push like you are 20 Points down!!  PUSH PUSH PUSH!!!!  VOTE JOE  He is not bat shit crazy. Thanks.  [@JoeBiden](https://twitter.com/JoeBiden)  [#JoeBiden](https://twitter.com/hashtag/JoeBiden?src=hashtag_click) | He's going to be back in Walter Reed or visiting Herman Cain by the weekend | You scared to debate again. CHICKEN🤡 |
| I want you and Kamala to know that I voted for both of you out of Florida this afternoon and I believe you can get our country back on track. | There is no one but Nancy Pelosi to thank for this. She refused to negotiate. | Tried everything you can but still can’t convince any of us |

**Exhibit 1:** Positive, Neutral and Negative replies (pre-processed) collected from both candidates’ tweets