Description:

The importance of Twitter as a communication tool is surging, especially during the election period. Politicians and political parties now maintain an active presence the micro-blogging platform. Election prediction using Twitter data is becoming a popular research area, where we can use the sentiments of citizens to estimate the performance of candidates in the general election.

In the 2020 election, we downloaded a dataset that includes the tweets related to election since Oct 15th. It uses keywords to collect Twitter data, such as the two main political parties, the Democratic and Republican parties, and the names of their candidates Donald Trump and Joe Biden. The dataset also provides location features so that we can get the user's city/state information. However, not every user has provided their location information. We believe that through user location information, we can have a more accurate understanding of public opinion. Predict election results more accurately under the winner-takes-all rule.

For user tweets, we need to perform sentiment analysis and use natural language processing (NLP) to observe the attitudes, opinions and emotions in the text. We used the Python TextBlob package to analyze the polarity of the dataset. Polarity analysis can determine the emotional attitude of the text author with respect to the topic being discussed. The TextBlob software package assigns negative and positive scores to words in the dictionary. These polarity scores are in the range of -1 to 1, where -1 represents extremely negative emotions, and 1 represents extremely positive emotions. A polarity score of 0 indicates neutrality.

We can use these scores to quantify the emotion of the text to get the user's attitude. In our experiment, we believe that when the score of tweet is positive, the user will support the mentioned political party or candidate, and vice versa. We analyzed the polarity scores of all tweets in the dataset to determine the voters supported by users. And using location feature, we listed the sentiments of users from the 10 most populous states in the United States in towards the two presidential candidates to assess their public opinion levels in these states. Finally, we use these sentiment scores of the two candidates to predict the winner in the election.

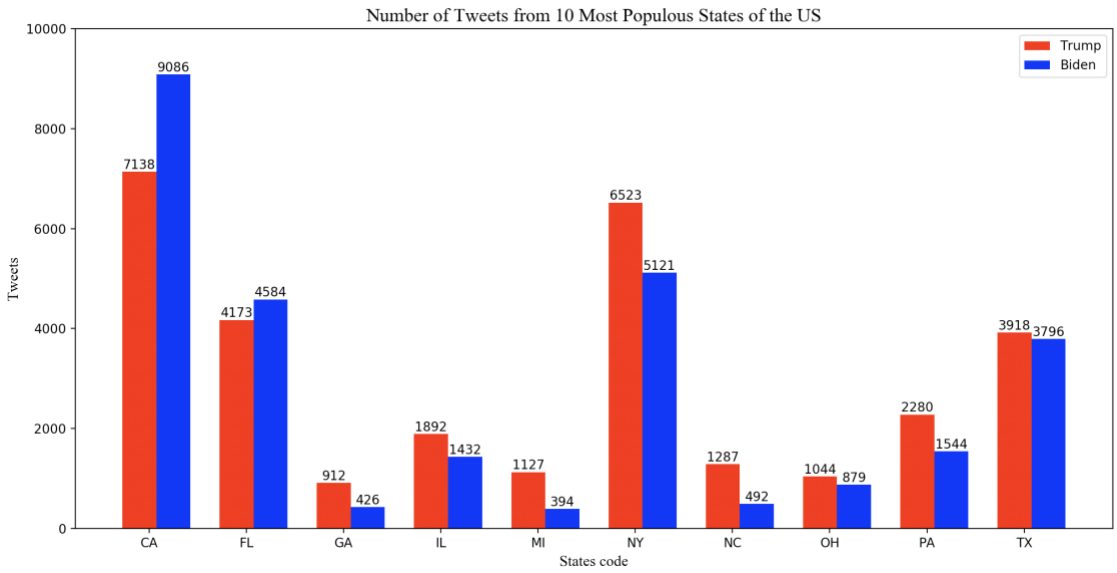

Figure.1 Number of Tweets from 10 Most Populous States of the US

Experiment results:

1) Sentiments results of users in 10 most populous states

Since it is not mandatory for Twitter users to share their location information, many tweets do not contain user location information. This dataset collected 342,511 tweets of valid users from all 50 states in the United States, of which 126,324 had valid user location information. We listed tweets from the 10 most populous states in the United States. Table.1 presents these 10 states account for almost 54% of the U.S. population and more than 46% of all tweets in our dataset. Figure.1 shows the number of tweets collected from these 10 states in the United States, and sentiments in these tweets in towards the presidential candidates: Trump and Biden. We can observe that democratic strongholds in California, and republican in NY.

	Trump	Biden	Total
All tweets from 10 most populous states	30294(52.19%)	27754(47.81%)	58048

Table.1 Sentiments of users in 10 most populous states

2) Sentiments results of all users

The sentiments of all the tweets were analyzed. We observed that 54.6% users has a positive sentiments to Trump in table 2. In this view, we can predict the Trump has the bigger chance to win the election in 2020.

	Trump	Biden	Total
All tweets	187011(54.60%)	155500(45.40%)	342511

Table.2 Sentiments of all users

In the future, we will collect data for a longer period of time and keywords to ensure that a sufficiently large representative sample is collected from all states.