In the 2018 Brazilian presidential election, Brazil saw unprecedented levels of online chatter and search activity. For this project, I hope to conduct analysis on tweets and google search trends as factors of prediction for election results. I will break down my results between the primary and final election rounds. My goal is to be able to pull insights from data and see if mentions on social media/internet searches is predictive of election winners. If my results are conclusive, I hope to pull similar data next year for the next presidential election and see if the model can be predictive for the 2022 presidential election.

I intend to use data from Twitter and Google, limiting results geographically to Brazil and temporally to three weeks prior to each election (the dates between the first and second rounds of the election, 10.07.18 and 10.28.18). I will gather tweets and search queries that I will be able to analyze for mention of presidential candidates. I will use Twitter’s Advanced Search tools and Google Trends to locate the data. I will then have to scrape these websites using tools and methods taught in class in order to fully have access to all the data I will need to conduct the analysis.

As mentioned, I will be heavily relying on web scraping to bring together the data I will analyze. I will then employ analysis of content and have a breakdown of every mention of the candidates in tweets and google searches. I will have two different sets of results: the first which will include the mentions received by 13 candidates who ran for President in 2018 which I will compare to the vote share received by each candidate.

I will create different graphs to illustrate the amount of mentions received by each candidate as well as the vote share in order to illustrate a conclusion on whether they are predictive or not for the first round. I will then repeat the same steps only for the two candidates of the second round of the election. I hope to then not only be able to see the effect of only two candidates running against each other on levels of chatter but also ultimately decide whether this chatter can be predictive of the full elections results. I will also be analyzing to what extent the chatter before the first round is indicative of the final round election results.

Success means my model will be replicable and I will be able to ascertain the predictive value of social chatter and searches in the 2018 election. Even if my project does not yield a result indicating that chatter was predictive in 2018, my goal is to have an accurate model that can work for analyzing chatter in any election, which I hope to use in 2022 to predict the outcome of the next presidential election.