# Quantifying a Candidate's Performance in a Political Debate


## Abstract

Political debates play an integral role in American politics. However, each party often spins the debate in order to crown their own candidate the victor of the verbal sparse. NLP provides a way to examine two candidates' performances through a more quantitative and objective lens. In this project, different NLP tools are employed to better determine which candidate "won" the presidential debates of 2020. The factors measured include a candidate's composure (Did their sentiment stay centered and positive?), objectivity (How truthful and objective were their claims?), consistency (How similar is this to their past statements?), and effective communication (Did they successfully convey their plans for presidency?). In order to answer these questions, the text of the debates was analyzed with TextBlob's sentiment analysis, NLTK's part of speech tagging, and Gensim's Doc2Vec model. While many of the results turned out to be similar between candidates, this work is an important first step towards quantifying a candidate's performance. All of these factors play an important role in deciding who did better in a debate, and by using these models, the general public can get a better understanding of each candidate's performance and decide who they think the true winner is. In the future, additional elements as well as more nuanced methods may provide an even better understanding of this topic.


## Introduction

In a day of ever changing political climate filled with online discourse and twisted news stories, political debates serve as an invaluable way for voters to get to know candidates and hear their stances straight from the source. With the upcoming presidential election, now more than

ever these debates serve as important displays of competence to the public. My family loves to put on presidential debates, but the real discourse comes after, when my father and I argue about who won. However, it's becoming increasingly hard to truly determine who "won" a presidential debate. Many voters watch with a jaded lens, simply declaring the winner to be the candidate from their political party. News channels are also not helping the problem. They fall into the same trap as viewers, spinning the debate to benefit whichever candidate agrees with their politics. In a world where there is rarely an objective news story, how can we more effectively quantify who won a debate, political party aside? In this project, I am exploring ways to quantify who won a political debate using NLP because news sources often twist the narrative and have subjective commentary. This will allow the public to better evaluate two political candidates and may inform their voting decision.

**Literature Review**

While there is no shortage of NLP research on the topic of politics, the vast majority of this focuses on the online discourse around politics instead of political statements themselves. Twitter and other social media are a gold mine of public reaction to current events. However, analyzing actual political texts and speeches is an undervalued area of study. By using NLP to analyze the words of those making political decisions, we can better understand who is running these establishments and how they are acting. In a study done in 2020, Chua Chin Hon used Hugging Face to measure the sentiment of political speeches in Singapore about COVID-19 (Hon). He paired shifts in sentiments with the topics being discussed to better understand the discourse at the time. In another study, researchers used NLP to analyze each country's statements in the UN General Debate (Baturo et al). They were able to identify the main topics

discussed in the United Nations' agenda setting and hoped that by doing so would make the process more transparent to the public. These are two brilliant examples of using NLP to directly analyze political discourse. In my case, I hoped to apply these same principles not to speeches or documents, but instead to presidential debates.

But how do we determine who "wins" a presidential debate? What factors sway the public with a better performance? Joe Pierre suggests that elements such as persistence, aggression, and strength may all play a role (Pierre). Other qualities might be the ability to keep composure and their effectiveness at conveying policy (Jade). In the end, four factors came to light. The first of these has to do with self-control: Did the candidate stay composed? In a setting where emotions are running high, personal attacks being spewed, and the whole country is watching, is the candidate able to stay cool, calm and collected? This is imperative to their debate performance and also an important quality for any politician to have. The second factor related to honesty: Is the candidate being objective and factual? Truthfulness is a huge part of political discourse, and often channels do fact-checking to show a candidate's honesty or lack thereof. Another aspect I wanted to measure was a candidate's ability to convey their policy: Did they explain the actions they would implement as president? Many minutes of a debate are spent dancing around questions and avoiding committing to any real answers. If candidates are able to successfully lay out their plans for the job, this would give them a better debate performance. Finally, I wanted to look at how a candidate's performance aligns with their past: Are these statements similar to their previous ones? Politicians are often criticized for flip-flopping on hot topic issues, so comparing to previous debates allows a better understanding of their opinions through time.

**Methodology/Dataset**

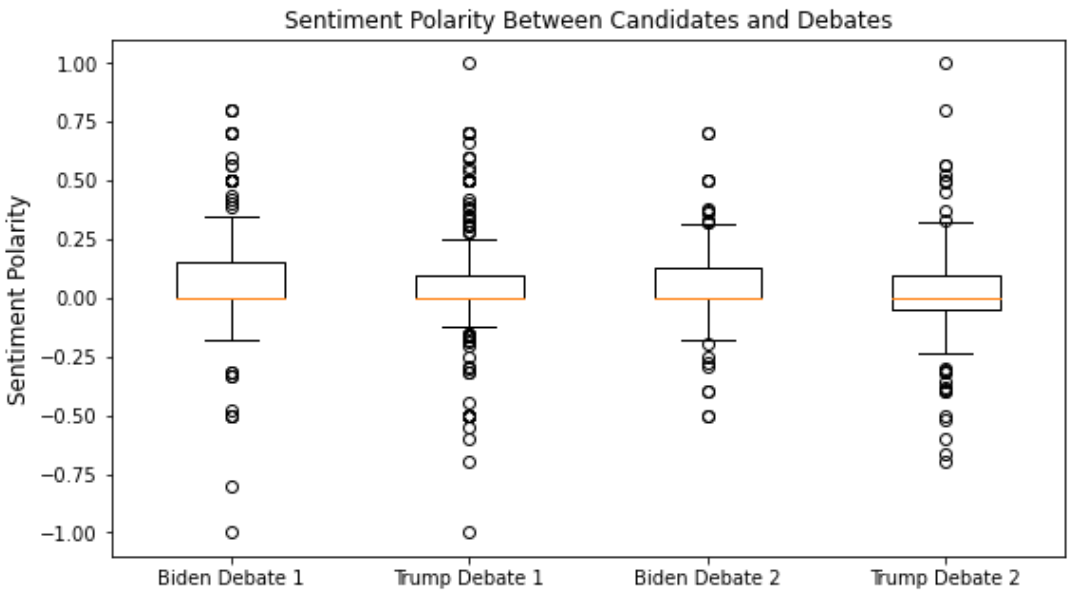
This dataset I worked with, found on Kaggle, contains the text of important televised events leading up to the 2020 election ("US Election 2020 - Presidential Debates"). It has both presidential debates, the vice presidential debate, Joe Biden's town hall, and Donald Trump's town hall. For each of these files, there is a text field with the sentences spoken, a speaker field which denotes who said the statement, and a timestamp which represents when the statement began. To preprocess the data, I used gensim's string preprocessor and stop words in addition to some custom ones. I also lemmatized using NLTK. This provided the tokens used for the majority of analysis. Calculating the time in seconds that each statement lasted was also a part of my preliminary analysis. However when I started trying to do this I noticed some problems with the timestamp field of the data. Partway through the debates, the clock would restart and begin at zero again. This threw a small wrench in my calculations, but I eventually fixed this problem. Although, I did have to go actually watch parts of the debate and record the length of a couple statements since those times were lost when the clock restarted. Eventually, I had the statements, their tokens, and how long in seconds the statements lasted.

After some preliminary EDA into things like statement length and time comparisons, I moved on to analyzing my four questions quantitatively. For the first question about composure, I decided sentiment analysis could be an insightful tool. By seeing how positive or negative a speaker's statements were, you could see if they went on heated tangents or stuck to being more composed and optimistic. Utilizing the pretrained TextBlob model, I calculated the sentiment polarity of each statement from the candidates. I then compared these distributions to each other to see how the candidate's discourse compared. Theoretically, the candidate with a more contained range and positive median would be better in this regard. Sentiment analysis also

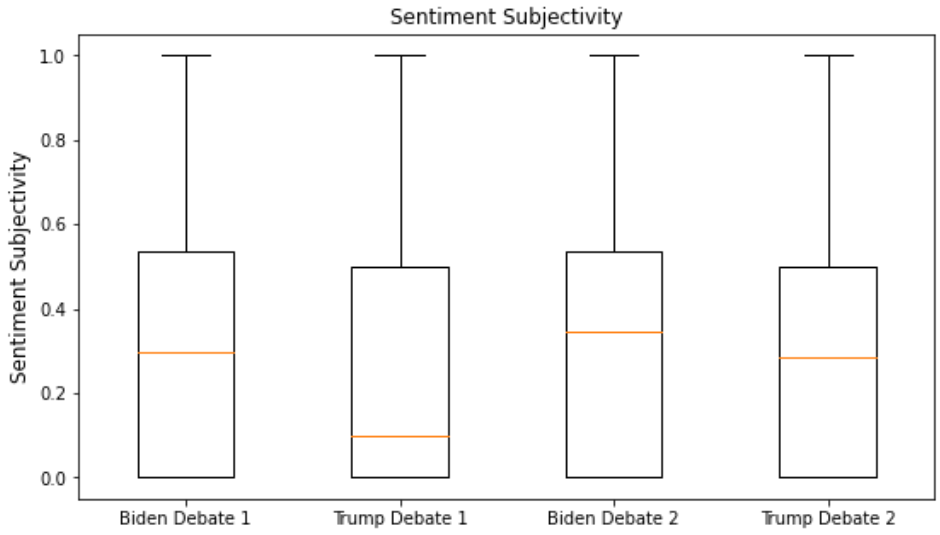
proved useful for the second factor of objectivity. TextBlob not only has sentiment polarity but also subjectivity. I completed the same process as before but this time looking at how objective a candidate's statements were. This obviously does not show if their claims were true, but it can still give us an idea of if they were backing up their arguments with evidence or just making outlandish statements. For the third factor, conveying policy, I decided to use part of speech tagging. I figured that if a candidate is truly answering questions about what they would do in office, their speech would have more verbs compared to someone dancing around questions. Hence, I used NLTK's POS tagging to detect verbs in a statement, and then found the ratio of verbs to total tokens, with a higher ratio indicating a better performance. Finally, for determining similarity to past performances I utilized gensim's Doc2Vec to cluster all the data I had. Theoretically, this model would consist of all of the candidates' public appearances in the past and the current debates would then be compared to this. However, since the debates and town halls are the data I already had, I decided to just cluster those.

**Results**

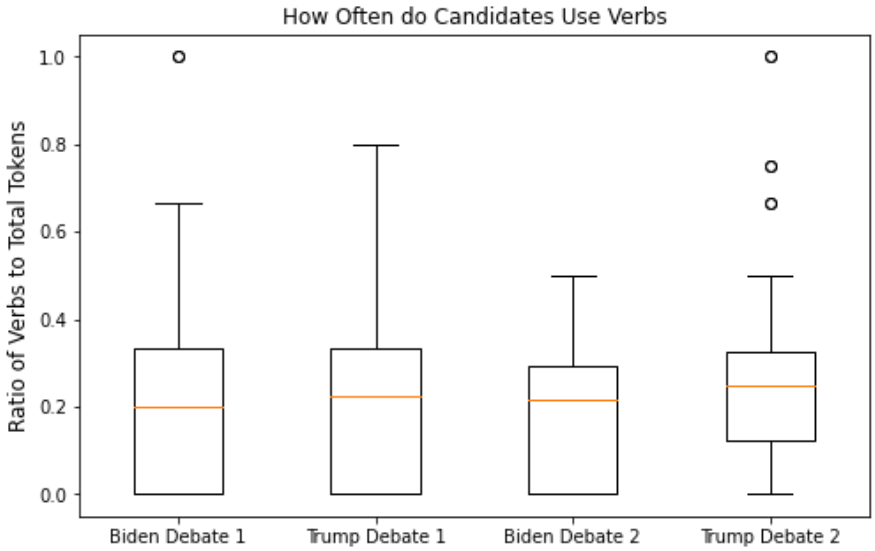
For the first question of a candidate's composure, I plotted the sentiment polarity of each candidate during the two debates, which can be seen below. Sentiment polarity measures how positive or negative their speech was, and ranges from -1 (the most negative) to 1 (the most positive). As you can see, they all have comparable medians across the board. Trump does tend to have a few more outliers compared to Biden for each debate. It seems that Biden in the second debate was the most consistent in his tone and didn't have any extreme positive or negative statements.



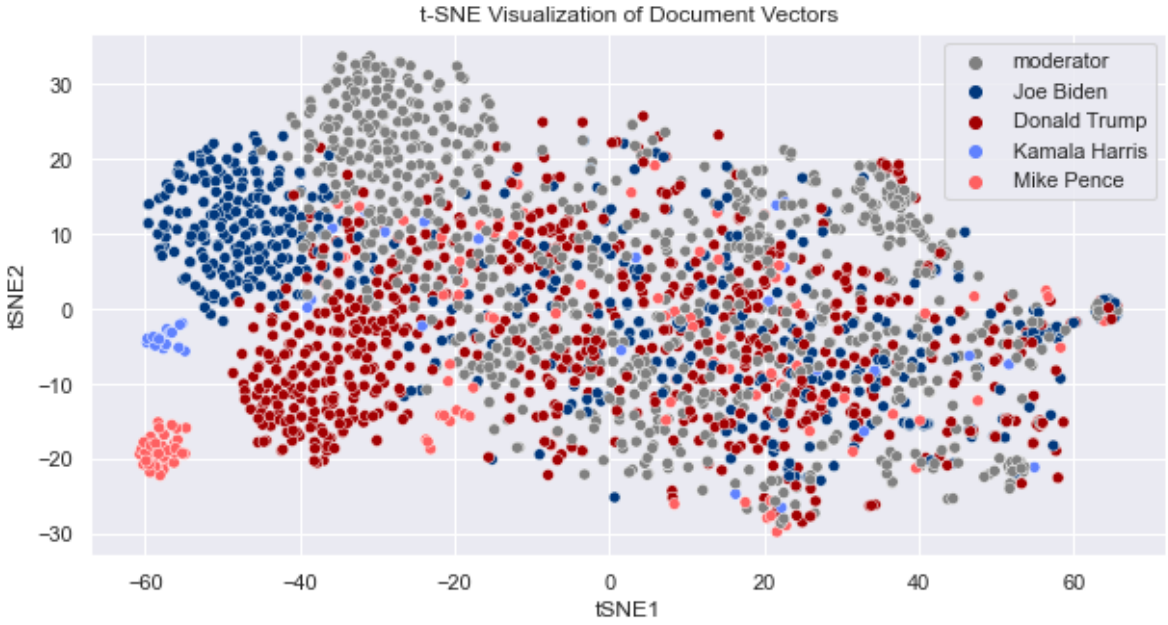

To address the second aspect, objectivity, I made a similar plot to the one above. Sentiment subjectivity ranges from 0, the most objective, to 1, the most subjective. Once again the plots are all very similar, much more so than I was expecting. It also came as a surprise that Trump's Debate 1 is the one with the lowest median subjectivity, meaning it was the most objective. Trump often comes under fire for spewing lies and not having an objective point of view, so this was a bit shocking. This is a case of when we have to remember not to just trust our numbers but to contextualize all findings.




To measure how well a candidate is conveying their plan of action, I compared the ratio of verbs to non-verbs in each of their sections of text. The plot of these ratios can be seen below. Once again these medians prove to be very similar. Trump seems to have a bit of a wider range, and his performance in the last debate especially stands out as being centered more around a higher verb ratio.



Finally, to look at the similarity between candidates' speech, I clustered the documents using gensim's Doc2Vec model and tSNE. Below, you can see that while there is a good bit of mixing, there is a clear group for each presidential candidate and their vice presidential candidate. This Doc2Vec model can then be used for future debates, using the similarity feature to see which past statement most closely corresponds to new ones.




Since there was not a definitive difference between candidates in any of the aspects I analyzed, I don't think there can be a clear "winner" declared. However, the analysis did provide new insights into the debates and gave a quantitative measure of qualities that are important for a presidential candidate.

**Discussion**

Overall, these four veins of analysis give us a new way to analyze the performance of presidential candidates in a debate. They help tie quantitative scores to a candidate's debate performance which will be helpful for objectively judging a possible politician. Things like

consistency, sentiment, and verbiage can be better judged by the public viewers. In the future, combining these factors and weighing them in a final model may give us a better understanding of who “won” a debate.

However, there is still a long way to go in NLP analysis of these debates and many shortcomings of the methods I used. For example, maybe having very positive and negative statements doesn’t come off as uncomposed, but is actually favorable in employing pathological arguments. Additionally, measuring subjectivity doesn’t tell you if the statements a candidate makes are true, but more so how they come across. In this way many of my methods grossly oversimplify complicated aspects of a politician's performance.

When compared to the studies analyzed in the literature review, I could have done more deep analysis on both sentiment and finding key topics. Finding key terms proved to be a challenge for me, but possibly by using a more complicated model like Baturo et al., I could find the most important policy issues. This could then be combined with a sentiment analysis like Hon did to compare sentiments of the two politicians while they spoke on certain subjects. While both of the examined studies dove deep into one quantitative measure, I chose breadth over depth, looking on the surface at four different factors to quantify performance.

Additionally, there are many more aspects that could be incorporated to better declare a winner. A big aspect that I think was missing from this data set is the crowd reaction. Many times the atmosphere of the event and the cheers and boos from the audience largely represent how a candidate is doing. Quick quips followed by big applause can show who is really dominating the event. Unfortunately, this data set did not have any audience reaction to the candidates speaking. In the future I definitely think this is something that should be included.

Other components such as interruptions, eloquence, and insults may also play a role and could be incorporated.

This process is complicated further by the fact that all of these “objective” qualities are interpreted subjectively. Different people have different priorities when evaluating a possible politician. For example, I prefer when a president is calm and collected, but other people may find aggression and fire to be more desirable. I appreciate when politicians recognize change in their constituents and reflect that in a change in their own opinion, however many people want their politicians to stay consistent throughout their careers. Hence, even if we quantify all of these different factors, it is still hard to declare an objective “winner,” as people's preference regarding these factors is distinct.

Overall, this topic is much more nuanced than it seems at first glance. While I was able to obtain some quantitative measurements, they didn’t prove extremely insightful and are complicated further by how the public would interpret them. In the future, I think that incorporating more factors and putting them into a final model would provide a better result. My ultimate end goal would be some sort of webpage for the public to examine the statistics of a recent presidential debate. Users could also interact with it, changing how important different characteristics are so the page can relay who performed better in those areas. For example, if having a positive outlook is very important to a user, then it would weigh sentiment analysis as a more important factor. The code I have set up now will provide an interesting comparison for the 2024 election debates, which will likely be a rematch between Donald Trump and Joe Biden. I hope to continue to hone my process using the new data that those debates will elicit.

**Conclusion**

Overall, I found that this project raised more questions than answers! My goal was to quantify a candidate's performance in a debate so that the public could better judge who won. I was successfully able to identify four important factors of a political candidate's performance and have a way to quantify those factors. However, my methods left me wanting to find more precise ways to measure these aspects and find even more factors that might play a role. This would allow a more holistic approach to identifying a winner and give the public a better sense of the debate.

## References


Baturo, Alexander, et al. "What Drives the International Development Agenda? An NLP Analysis of the United Nations General Debate 1970-2016." *Cornell University Library*, 9 Aug 2017. https://medium.com/bigger-picture/what-determines-who-wins-a-political-debate-2bd1adeb2aaf

Hon, Chua Chin. "Sentiment Analysis Of Political Speeches Using Hugging Face's Pipeline Feature." *Medium.* https://towardsdatascience.com/sentiment-analysis-of-political-speeches-using-hugging-faces-pipeline-feature-3109c121d351

Jade, Lady. "What Determines Who 'Wins' a Political Debate?" *Medium.* https://medium.com/bigger-picture/what-determines-who-wins-a-political-debate-2bd1adeb2aaf

Pierre, Joe. "Scoring the Presidential Debates: How Do We Decide Who Wins?" *Psychology Today.* 26 Sep 2016. https://www.psychologytoday.com/us/blog/psych-unseen/201609/scoring-the-presidential-debates-how-do-we-decide-who-wins

"US Election 2020 - Presidential Debates" https://www.kaggle.com/datasets/headsortails/us-election-2020-presidential-debates