**Overview**

Sentiment Analysis (SA), through techniques like Natural Language Processing (NLP), is now a widely used data technique. Companies and individuals reply on SA for powerful insights and apply affective subjective information into their growing text-based data. SA is often applied to add context to the customer’s voice during items such as reviews, surveys, and social media posts.

NLP techniques are now being applied to live communications. In areas like health care (Siwicki, 2021), NLP mines text in documents and transcripts of doctor’s notes to identify facts, relationships and assertions which would not have been seen before. The extracted data can be further analyzed for trends and gain the impact of social, economic, and environmental factors leading to the individual health and quality of life.

In another example, companies and states are running the transcripts of legal cases through NLP engines to gain close to the same insights as in the medical example. Parole hearing transcripts (Hong & Voss, 2021) can be searched to answer key factors such as when the last disciplinary infraction was and what was the cause. These models can speed up and raise the accuracy of important legal proceedings.

The question we are discussing fits between those two models, can we run SA and NLP over the transcript of a formal written speech by the United States President and gain useful information. Next, can that information be used to track changes in their job performance as seen through the Gallop polls? Is it only specific ‘Defining Presidential Moments’ which can correlate between the speech and the job approval numbers? Is there other factors we can have associations or links to approval number changes?

Schubert, Stewart, and Curran (Schubert, Stewart, & Curran, 2002) created a study in 2002 to judge the effects and rallies of nationalism after a major event driven by presidential speeches, specifically the effects of George W. Bush in the days after the 11 September attack. Their results proved a dramatic jump in approval after his speech to congress the day of the attack. Later, in this analysis, we will see if our data backs up these results.

Putting presidential speeches through SA models was also performed by John Paul Miranda and Rex Bringula in 2020 (Miranda & Bringula, 2021) and by Renalyn Banguis-Bantawig in 2018 (Banguis-Bantawig, 2019). Miranda and Bringula used the State of Nation Address (SONA) of the past 13 Philippine presidents the team analyzed the general sentiment of the addresses during specific periods of Philippine governments. Their findings showed the positive and negative sentiments followed the general state of the period. For instance, the speeches with the lowest sentiment scores were during the martial law periods around 1974. Banguis-Bantawig used 54 speeches of selected Asian Presidents to analyze their methods of reaching across the podium to sway the opinions of the audience. The results of both studies prove the importance of a presidential “speech” as a venue to discuss and demonstrate the nation’s state and direction.

Matthew Eshbaugh-Soha (Eshbaugh-Soha, The Impact of Presidential Speeches on the Bureaucracy, 2008) states “...*presidential speeches* are an effective means of influence over bureaucratic activity, and that this influence is contingent on the direction of the president’s policy signals” (p. 129). Positive signals in the speech reinforce the mission and motivate bureaucrats to do more. In contrast, negative signals have little to no impact on the same mission or motivation. He does not go into the effect of the third aspect of SA, the neutral signal. We will need to delve into its effects.

Model

Sentiment Analysis works by applying a score to specific words. The scores are complied to give us general direction of the text.

In this paper we will use the VADER (for Valence Aware Dictionary for Sentiment Reasoning) (Hutto & Glibert, 2014) through TextBlob. While the VADER model is specifically geared to social media text, we chose this method as a speech, while written, is still created to be spoken. Social Media is also another form of same process, it’s written but tends to the less formal aspects and structures of literature.

TextBlob (Loria, 2020) gives us 5 key values per text set: Positive, Negative, Neutral, Polarity, and Subjectivity. Positive, Neutral and Negative scores are the percentage ratio of words in the text which fall into those 3 buckets; as an example, a text can have 20% Positive, 77% Neutral, and 3% Negative words leading us to summarize the text as more positive than negative. Since we are analyzing speech, we will always have a neutral value as it would be very difficult to effectively communicate while only using positive and negative words. We can also see the effects of a high value in neutral words on the sentiment and the job scores. Polarity is a single value score which lies between [-1,1] where a -1 indicates a highly negative sentiment and a 1 indicates a highly positive sentiment. Negation words reverse the polarity. Subjectivity is a single value score which lies between [0,1]. The closer the number is to zero the more subjective the text. In reverse, the closer to 1, the more objective the text is.

Data and Data Preparation

Gallup, INC is an American analytics and advisory company which operates from Washington D.C. Founded in 1935, the company is best known for its public opinion polls conducted worldwide and has become one of the accepted sources of US politics public opinion. Since 1935, the company has performed the same style of opinion poll to generate the US public’s opinion on how well the President is doing in his job.

For each President, the poll shows the percentage from 100 of the those who Approved, Disapproved, or were unsure. A percentage change formula was applied for each of those columns to capture the change in sentiment from the previous poll. The method allows us to see how much change, if any, occurred between the polls.

The poll, however, was not run on a regular basis until recent years. While we have almost weekly poll numbers for Presidents George W Bush and Barrack Obama, earlier Presidents, such as Dwight Eisenhower we have large gaps between the poll dates. There was no clean method to programmatically match the poll numbers and their changes for each speech; a manual matching process was followed. Where possible, a speech was matched up with the corresponding poll numbers of the week following the speech. This allowed for the speech to have time to be heard and a poll organized to capture the public mood.

Through the University of California, Santa Barbara, we received the combined datasets for all presidents Truman through Trump (Peters, 2021).

# Bibliography

Banguis-Bantawig, R. (2019). The role of discourse markers in the speeches of selected Asian Presidents. *Heliyon*, e01298.

Bermingham, A., & Smeaton, A. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results. *SAAIP* (pp. 2-10). Chiang Mai: IJCNLP.

Ceron, A., Curini, L., & Lacus, S. (2015). Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence Fromthe United States and Italy. *Social Science Computer Review, 33*(1).

Dilin Liu, L. L. (2018). *The appeal to political sentiment: An analysis of Donald Trump’s and Hillary Clinton’s speech themes and discourse strategies in the 2016 US presidential election.* Amsterdam: Elsevier.

Eshbaugh-Soha, M. (2008). The Impact of Presidential Speeches on the Bureaucracy. *SOCIAL SCIENCE QUARTERLY, Volume 89, Number 1,*, 116-132.

Eshbaugh-Soha, M. (2010). The Politics of Presidential Speeches. *Congress & the Presidency, 37*, 1-21.

Haselmayer, M., & Jenny, M. (2014). Sentiment analysis of political communication: combining a dictionary approach with crowdcoding. *EPSA Confrence.* Edinburgh: Springer Link.

Hong, J., & Voss, C. (2021, 05 03). *HAI Stanford University*. Retrieved from Is Natural Language Processing Ready to Take on Legal Hearings?: https://hai.stanford.edu/news/natural-language-processing-ready-take-legal-hearings

Hutto, C., & Glibert, E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.* Palo Alto, California: Association for the Advancement of Artificial Intelligence.

Lauderdale, B., & Herzog, A. (2016). Measuring Political Positions from Legislative Speech. *Political Analysis, 24*(3), 374-394.

Loria, S. (2020, 04 26). *TextBlob*. Retrieved from TextBlob: Simplified Text Processing: https://textblob.readthedocs.io/en/dev/

Lowe, W. (2008). Understanding Wordscores. *Political Analysis Vol 16*, pp. 356-371.

Miranda, J., & Bringula, R. (2021). Exploring Philippine Presidents’ speeches: A sentiment analysis and topic modeling approach. *Cogent Social Sciences*, 7: 1932030.

Peters, G. (2021). Presidential Job Approval. *American Presidency Project.* UC Santa Barbara. Retrieved from https://www.presidency.ucsb.edu/statistics/data/presidential-job-approval

Schubert, J., Stewart, P., & Curran, M. (2002). A Defining Presidential Moment: 9/11 and the Rally Effect. *Political Psychology , Sep., 2002, Vol. 23, No. 3, Special Issue: 9/11 and Its Aftermath: Perspectives from Political Psychology*, 559-583.

Siwicki, B. (2021, 02 19). *Using natural language processing to unlock SDOH in unstructured EHR data*. Retrieved from https://www.healthcareitnews.com/: https://www.healthcareitnews.com/news/using-natural-language-processing-unlock-sdoh-unstructured-ehr-data