Introduction

This project will analyze Obama's speeches during his two consecutive presidential terms from 2009 to 2017 using multiple natural language processing techniques. The goal of this study is to apply topic modelling and sentiment analysis on Obama's speeches , also to understand how external factors influence the frequency of different topics in Obama's speeches.

Data Pre-processing

The content and URL of each speech were gathered using BeautifulSoup and the Python packages urllib. I retrieved the URL to extract the data from the "href" tags. After obtaining the URLs, duplicate links were removed, and prefix all of the URLs that begin with "speeches" and end with "htm" were preserved with "https://www.americanrhetoric/". All speech titles follow the same format and have the color formate "#000048". The dates were recorded in <font, size='4', face='Tahoma'>, and contents were stored in <font, face='Verdana', size='2'>. I could retrieve all the information I need by looping over each URL. The data was then saved into "obama speech.csv" CSV file for further usage.

During data pre-processing, I clean the text using a variety of NLTK techniques. I modified the text's case to lowercase and got rid of all the punctuation. I then removed the unnecessary information by combining the package's stopword with other stop word extensions. The word is then converted back into its original lemma form for nouns and verbs using the wordNetLemmatizer. In order to train the model in the next step, the data frame was then saved as "Obama clean.csv".

Topic Modelling

Gensim LDA, Sk-Learn LDA, K-means, and NMF—are used to assess the subjects of speeches. Its objective is to abstract the subjects found in a group of papers that have a manageable number of topics.

The optimal number of topics using the Gensim-LDA and Sk-Learn models is 11, according to the coherence score function. Since the results were more effective and understandable than those obtained using the initial topic count of 15.

I next applied NMF and K-means models into practice. K-means splited 428 speeches into 12 clusters after the number of topics was modified using the elbow approach. Lastly is NMF, the distribution of the features in each topic was calculated, and the top 10 features were then extracted to determine the topic's content. 11 subjects were found as the results, which are comparable to earlier findings.

Sentiment Analysis

I utilized sentiment analysis to determine whether his speeches were more positive, negative, or neutral after categorizing all the speeches according to topics. Gensim and Textblob were used

for the analysis. I divided the findings into 5 categories (strong negative, negative, neutral, positive, and strong positive) based on the polarity in TextBlob and the compound positivity and negativity scores in the Gensim model.

Summary

From the topic modelling output, I finalized the topics into nine large categories: 1. Immigration and international trades, 2. Economy and job opportunities, 3. Freedom and democracy, 4. Obama's re-election, 5. Nuclear War and Military force in Iraq, Iran,Isil and Syria. 6. Health care, 7. Crime and Violent, 8. Intelligence security and Surveillience 9. Education and family safety.

All of Obama's actions and statements had to do with jobs, such as American Recovery and Reinvestment Act in 2009, addressing pay disparities between men and women in 2009, and introducing the American Jobs Act in 2011. These actions and statements had a significant impact on the unemployment rate and increased public optimism about the economy and democracy. Obama delivered more speeches on health care between 2010 and 2013 as a result of the actions of the Health Care Reform Bill in 2010, the Affordable Care Act in 2012 to 2013, and the emergence of the Ebola virus. Compare this with the expansion of health benefits coverage in the US, where more individuals now have access to affordable health insurance as a result of the Affordable Care Act.

Throughout his tenure, Obama has delivered a number of speeches on education reform between 2009-2014, including those on student loans, veteran education reform, and the No Child Left Behind Law. The likelihood that Americans will pursue higher education has increased as a result of all of these factors. Obama's actions and comments advanced US educational progress and led to official school reform.

Obama offered the idea of a nuclear-free world after 2009 , also he frequently discussed nuclear security, counterterrorism, and promoting peace and lessening war between 2014 and 2017. During the same time period, military budget and the quantity of nuclear weapons also significantly fell. Obama's statements have undoubtedly had a big impact on the military and nuclear security of the United States.

Conclusion

I collected and cleaned the data for analysis using BeautifulSoup and NLTK. Then, to categorize the speeches, I used Gensim LDA, Sk-Learn LDA, K-Means, and NMF. Finally, Sentimental Analysis was applied to evaluate the sentiment of the speeches using Textblob and Gensim. In conclusion, 428 of Obama's talks may be grouped into about nine clusters. The majority of his speeches focus on democracy, the economy, health care, and war. Based on our test, Obama's remarks had a more positive sentiment. Along with Obama's increasingly optimistic economic statements, the US labor market and economy has been improving. Obama's remarks and actions led to considerable improvements in health care coverage as well.