**Trump Tweet Sentiment Analysis**

**Executive Summary/Abstract**

The purpose of this project was to perform topic analysis on Donald Trump's Facebook status updates to gain insights into the content that resonated with his audience. The content of Trump’s Facebook status updates was analyzed to create coherent topics using LDA topic modeling. The top 10 most frequent topics were analyzed to determine whether respondents reacted positively or negatively. Additionally, the topics that generated the most shares and reactions were analyzed to gain a better understanding of the content that was most effective in engaging Trump's audience on Facebook. Overall, the project provides valuable insights into the content that resonated with Trump's audience on Facebook and can be used to inform future social media strategies.

**Introduction/Background**

This project analyzes Donald Trump's Facebook status updates dataset to determine the sentiment and topics that generated the most engagement from his audience. The dataset includes various metrics such as the number of reactions (likes, loves, sads, and angrys), shares, and comments for each post. Additionally, the dataset includes the status message text and status type (photo, video, link, or status). LDA was used to extract coherent topics from Trump’s status updates so that various charts and analysis techniques could be applied to visualize the data and draw insights into the audience's reactions to Trump's posts.

**Methods and Results**

First, the data was loaded in, unnecessary columns were dropped, rows with na’s were dropped, and columns were converted to appropriate data types. The data was then plotted to showcase the distribution of status types as seen below:

Chart, histogram

Description automatically generated

A new dataframe was created using the date of publish as an index so that the reactions could be plotted over time which is posted below. The number of likes vastly outnumbered the other reactions.

Chart

Description automatically generated

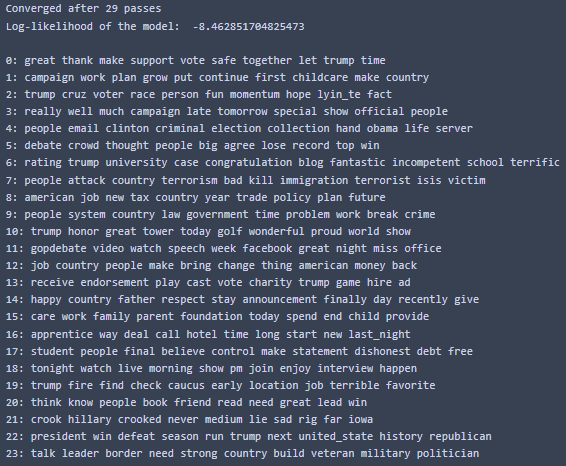
The next goal was to extract topics from the status updates but first, the data had to be preprocessed in order to maximize topic coherence. The content of the status updates had stop words removed, words were lemmatized, and bigrams and trigrams were created.

Once the data was processed, the first step of topic extraction was to determine a convergence baseline for the number of passes. Once the initial optimal passes parameter was determined, multiple models were created to test the optimal number of topics to maximize topic coherence. The results were plotted and analyzed.

Chart, line chart

Description automatically generated

The most notable coherence spikes can be seen at 8, 12, 20, and 24 topics. New optimal pass parameters were determined (convergence) for each number of topics and models were trained. The keywords of the topics were analyzed by hand to determine which ones had the most coherence to the human eye. The lower topic-count models appeared to group multiple topics within one which suggested more topics would be ideal. Ultimately, the 24-topic model was chosen to be used for further analysis. The top 10 keywords for each topic can be found below. Additionally, the frequency topic distribution can be found below the keywords.



Chart, histogram

Description automatically generated

The next goal was to determine whether those who interacted with Trump’s updates had a positive or negative reaction. The topic modelling results were incorporated into the working dataframe and a new dataframe that contained data for only the 10 most frequent topics was created. New columns were created combining the number of likes and loves into a ‘happy’ column and sad and angrys into an ‘upset’ column. This dataframe was then pivoted in order to create a heatmap. The heatmap showed that positive responses largely dwarfed negative responses in all of the top 10 topics. The most liked and disliked topic contained keywords that suggests it encapsulates status updates pertaining to Trump winning the presidential election.

Table

Description automatically generated with medium confidence

Lastly, the entirety of the topics were evaluated to determine the top 5 topics that were shared and the top 5 topics that were reacted to. These groups largely overlapped with 4 out of the 5 topics being found in both groups. Histograms detailing the topics’ engagement can be found below.

Chart, bar chart

Description automatically generatedText

Description automatically generated

Chart, bar chart

Description automatically generated Text

Description automatically generated

**Conclusions**

After analyzing the dataset, I’ve concluded that Trump’s facebook statuses are largely reacted to by people who like him. Topics that had negative connotations such as terrorism in topic 7 or Hillary Clinton in topic 21, had lower total reaction levels than topics that had positive connotations such as topic 22 (Trump winning the election). Despite its negative connotation, topic 7 (terrorism), was the second most shared topic. It also had the 3rd most positive reactions on average. However it did not make it into the top 5 reacted to topics. This suggests that although status updates regarding terrorism were less frequent than other topics, the response was zealous when they appeared.

One of the challenges was determining the optimal number of topics to model. I am happy with my results using 24 topics however I believe there is still room to fine tune the results for more optimal coherence.

Another challenge was determining how to handle how positive and negative reactions were calculated. Due to the nature of the data, the number of variables composing each measure were imbalanced. In order to maximize readability of the heat map, the number of laughing reactions were not included in the final positive score.

**Future Research Considerations**

In future research I would like to analyze the frequency of the topics over time. It is possible that analyzing the relationship to historical events could lend a hand in determining the theme of each topic.

**References**

Delpisheh, E. (2023, March 25) Donald Trumps Facebook Statuses [Dataset].