Sentiment Analysis on USA Presidential Debates of 2020

Prajakta Phadke (pphadke@iu.edu) Mark Doucette (madoucet@iu.edu)

# **Introduction**

# **Backgrounds**

# **Research question**

The paper aims to address the following questions using the data collected from twitter:

1. How were the United States Presidential election debates perceived by people?
2. What were the most popular topics that were talked about in the tweets?

# **Data**

The data for this study was collected from Twitter. Twitter data is ideal for this study because this data comprises of people’s own thoughts and opinions. The research questions aim to evaluate these opinions and study their views. There have been two presidential debates in United States for which the data is collected. Following steps were performed to collect the data:

1. Twitter data was queried using the twitter api and using a full archive search. The tweets were taken from 2020 and there are about 5000 tweets.
2. These tweets were collected in their extended format and some with just the tweet text when the extended format was not available. The hashtags used to query this data are #Debates2020 #Presidentialdebate, #2020debate.

[Mark – if you can add more points to how the data was filtered.]

# **Methods**

The aim of the paper is to evaluate people’s views on the presidential debate and understand the topics that were most talked about in these tweets. In this paper we plan to perform two main types of analyses, sentiment analysis and topic modeling. Below are the detailed steps performed to do so:

1. After the preprocessing of the tweets is done the first step was to determine the polarity or sentiment of each tweet. There are many python libraries that do this. I used the textblob library for the same since it is one of the well-known libraries for this task.
2. Once I got the sentiment of all the tweets (these had three values, positive, neutral and negative), I plotted histogram for the sentiments for the full data. This would show us the overall sentiment regarding the debates.
3. The second part was to perform topic modeling. This included two steps. The first was to plot a word cloud to see the kinds of things people are talking about. And the second part was to use Latent Dirichlet Allocation (LDA) for topic modeling
4. Just like sentiment analysis, we did the topic modelling steps mentioned above on the whole dataset.

The trends, outputs and results obtained from these methods are shown in the results section of the paper.

# **Results and Observations**

1. **Timeline of the tweets:**
2. **Sentiment Analysis**

1. **Topic Modelling:**
   1. **Word Cloud**

The definition of Word Cloud is that it is an image which is formed by words that are present in the textual data. The size of the words in the image is directly proportional to its frequency in the text. It can also be interpreted as more important the word, bigger it is in the cloud. The word cloud for the whole data tweets looks like:

[Insert Word cloud]

[Insert observations about the word cloud]

* 1. **Latent Dirichlet Allocation**

The LDA analysis for topic modelling is a statistical modeling technique used to identify some topics that occur in the data. Latent Dirichlet Allocation (LDA) uses a topic per document model and a words per topic model. They are together modelled as Dirichlet Distributions.

[Insert some analysis about the observed results]

1. Topic1
2. Topic 2
3. Topic 3
4. Topic 4
5. Topic 5
6. Topic 6
7. Topic 7
8. Topic 8
9. Topic 9
10. Topic 10

The dashboard for this topic modelling looks like:

[If we want to show the dashboard, but this is where I

Got the feedback that it was too small to see. This screenshot is from my previous paper as an example]

Chart, bubble chart

Description automatically generated

# **Conclusion**

# **References**

[1] Jennifer Andriot, Baekkwan Park, Peter Francia,” Sentiment Analysis of Democratic Presidential Primaries Debate Tweets Using Machine Learning Models”, 07//2020

[2] Brandon Joyce, Jing Deng,” Sentiment analysis of tweets for the 2016 US presidential election” 11/05/2017

[3] Malak A. Abdullah, Mirsad Hadzikadic, “Sentiment Analysis of Twitter Data: Emotions Revealed Regarding Donald Trump during the 2015-16 Primary Debates”, 11/2017

[4] Ussama Yaqub, Soon Ae Chun, Vijayalakshmi Atluri, “Sentiment based Analysis of Tweets during the US Presidential Elections”, 06/2017

[5] Pekka Isotalus, “Analyzing Presidential Debates”, 06/2011