Presidential Campaign Debate Text Analysis

This project involved analyzing 2016 US Presidential campaign debate speeches for the amount of 'negative words' they contained and the effect this had on their polls throughout the campaign trail. Further project instructions are listed below.

A news company has approached you as a data scientist to find out if you can identify the effect of 2016 election debates and speeches on the polls.
You will use the 2016 U.S. Presidential Campaign Texts and Polls dataset from Kaggle (https://www.kaggle.com/alandu20/2016-us-presidential-campaign-texts-and-polls). It contains the following:

Collection of data from the 2016 U.S. Presidential Election Campaign containing:

1. Transcripts of the three presidential debates, divided into separate Trump and Clinton text files
2. Transcripts of Trump's 64 speeches delivered after the RNC and Clinton's 35 speeches delivered after the DNC
3. Transcripts of select speeches delivered by candidates during the primary campaigns
4. USC Dornsife/LA Times Presidential Election Poll, with daily breakdown by voter groups
5. Five Thirty Eight Election Poll, containing daily data from numerous pollsters

Your submission should include:

1. A summary of your findings
2. Sources used.
3. Technology used (NEW!!!)
4. Method of discovery
5. Graphs/diagrams of results

Additional information for the project:

- Using the attached negative words list estimate the level of negativity in each speech. This is an estimate and does not take into account stemming and lemmatization. This is a simple statistical count.
- Using the dates of in the name of each speech, correlate with the poll numbers to determine the effect of the speech on the poll numbers.
- Explain the possible pitfalls of your process.

https://www.cs.uic.edu/~liub/FBS/Sentiment-Analysis-tutorial-AAAI-2011.pdf