

# Sentiment Analysis – Graphing Results And Conducting Analysis

*AP Research & Senior Research Project*

Ojas Chaturvedi

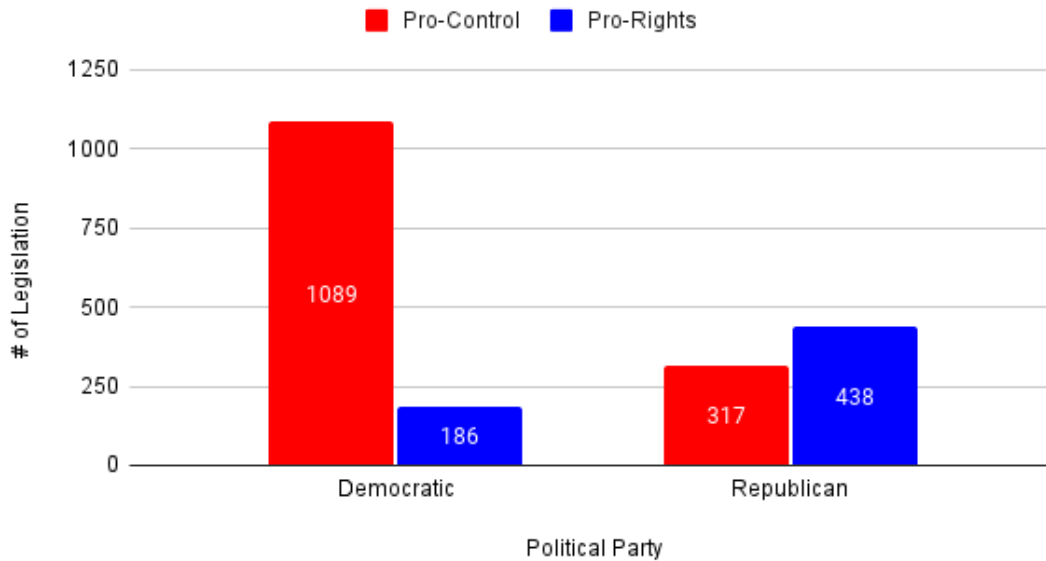
<https://github.com/ojas-chaturvedi>

---

This week was dedicated to first finishing up the results section of my project and then working on a practice presentation for rehearsals. As I mentioned in [last week's blog](#), there are 3 types of sentiment analysis models: Rule-based, Machine Learning, and Transformers. After some additional research, I decided to not go down the Machine Learning or Transformer model path. Both those types of models can have the highest accuracy, but only with pre-labeled data as a training guide for the models. The pre-labeled data would essentially teach the models how to correctly assign sentiment scores in alignment with the methods used in the training dataset. However, pre-labeled data needs to be extremely accurate, otherwise, it would introduce bias into the model and give extremely inaccurate results, which is why pre-labeled data is made by experts in the field who go into some training examples and classify sentiment by hand. I did try to find any training data possible, but couldn't find anything relating to congressional legislation, specifically regarding the gun control debate. Therefore, I decided to stick with Rule-based models, and more specifically VADER (Valence Aware Dictionary for sEntiment Reasoning), as it has pre-defined lexicons made for the general English language which I could fine-tune myself to increase the accuracy. Fine-tuning the lexicon involved going through the dictionary file, finding any words that would have a different sentiment in a legislative context compared to a general English context, and manually changing the sentiment values. For example, the word "vice" generally has a negative sentiment with the definition being "immoral or wicked behavior", but in the legislative context, "vice" was only present in "Vice President", which should have zero sentiments.

After fine-tuning the lexicon for the VADER model, I ran it for all the legislative texts and saved the results in CSV files found in the [project GitHub repository](#). The next step with the results is to start generating the graphs to see trends and visualize the data to determine significance. The first graph I created was a bar chart with the 2 political parties and legislation classification. I wanted to see what side each political party leads to. Below is the graph:

### # of pro-control and pro-rights legislation by political party

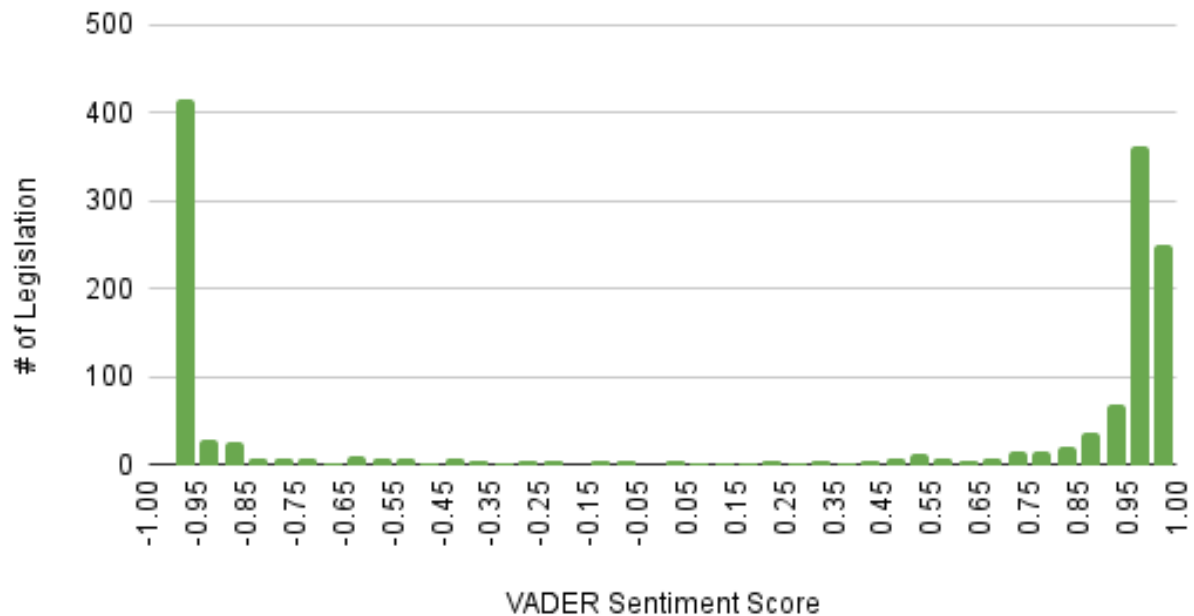


*Figure 1: Comparative Bar Chart of Legislative Trends: This graph presents a quantitative analysis of pro-control and pro-rights legislation as supported by the Democratic and Republican parties, revealing the contrasting priorities and policy focus within American political dynamics.*

As seen in the bar chart, the Democratic party has ~5.85 times more pro-control legislation than pro-rights, showing that the Democratic party leans heavily towards the pro-control side of the gun control debate. Meanwhile, the Republican party is more split, having ~1.38 more pro-rights than pro-control legislation, proving that the Republican party is more split for this issue but does still lean more toward the pro-rights side.

I wanted to see the overall distribution of legislation based on sentiments to visualize the polarity of the overall legislation. Therefore, I created a histogram with a 0.05 step of all the sentiments of congressional legislation:

## Sentiment Distribution of Congressional Gun Legislation

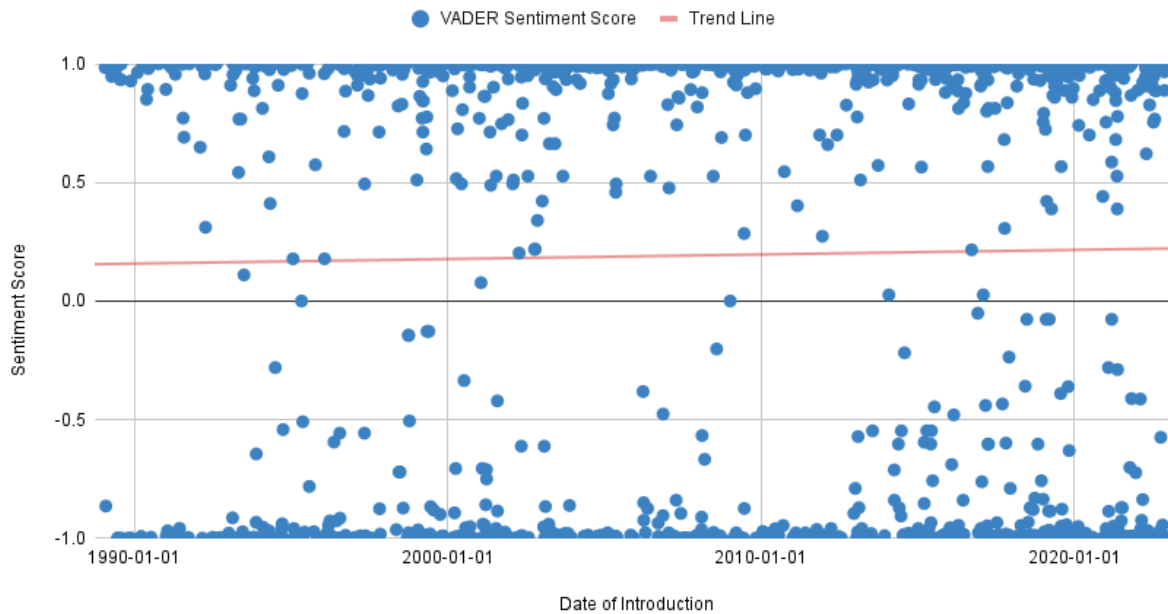


*Figure 2: Histogram of Congressional Gun Legislation Sentiment: This histogram provides a distribution of sentiment scores across all gun-related legislation, as measured by VADER sentiment analysis, to reflect the general emotional tone within legislative language.*

As seen in the histogram, congressional gun legislation is extremely polarized. There is barely any legislation in the neutral area, which is a sentiment value between -0.05 and 0.05. Instead, the histogram has a bimodal shape, with the 2 peaks being at the most positive and most negative scores.

The next graph I created was a time graph plotted with sentiment scores to see how the overall legislation sentiment changed over time. Below is the graph:

### Change in Sentiment of Congressional Gun Legislation Over Time



*Figure 3: Temporal Analysis of Sentiment in Congressional Gun Legislation: This scatter plot with a trend line demonstrates the fluctuation in sentiment scores of gun legislation over three decades, as evaluated by the VADER sentiment analysis, indicating the evolving emotional tone in legislative language.*

As you can see by the time graph, the sentiments of the legislation are very polarized, with a lot of legislative texts having values near 1.0 and -1.0. I plotted a general trend line using the data to get an equation for how time affects the sentiment and received this equation:  $y = 0.00000529x + -0.0161$  where  $y$  is the sentiment score and  $x$  is the date of introduction. The equation shows that as time goes on, the sentiment in legislation does become more positive but in extremely small increments.

In the upcoming week, I plan to continue writing my research paper for final submission on April 29th. Currently, all of my sections are in notes and not formal writing, so I will be spending a lot of time writing the final draft of my paper. Once I finish the methodology, results, and analysis sections, I will go back to my introduction/literature review section and continue improving the sections by reading more papers in the current field. Meanwhile, I will be completing my AP Research and internship tasks, as always.