

Analysis of Bias in The Political Speech of Gun Control Based On TF-IDF

Author: Heeseo (Kay) Lee

Korea International School, Pangyo, Republic of Korea

hlee24@student.kis.or.kr

ABSTRACT

Gun violence is an issue that has been widely debated in the field of politics, especially in the United States, where gun violence still remains a large problem. In the political field, this issue has sparked various debates on what steps the U.S. government should take to either stricken or reduce laws regarding gun control. In the modern world, the words of these politicians are all the more influential to citizens, and should therefore be duly criticized, especially on a matter that deals with life and death, such as gun control. This essay analyzes the attitudes of politicians in the United States regarding gun control through their many speeches by utilizing both the TF-IDF data sorting algorithm and the human mind. Known for its efficiency, the TF-IDF algorithm is used for a variety of purposes, such as data-mining, text analysis, and language processing. This essay utilizes this algorithm to effectively sort and rank the relevance of terms in a corpus containing 20 speeches on gun control given by American politicians, therefore allowing for an efficient analysis of the main attitudes of political figures in the U.S. towards gun violence. Results conclude a variety of opinions on the topic.

Keywords: Attitude, Gun control, Gun violence, Politicians, TF-IDF, Datamining

1. INTRODUCTION

1.1 History of Gun Laws

Gun violence, and subsequently, gun control, are two subjects that have undergone large amounts of debate across the globe, but especially in the United States. More than 500 people die of violence due to firearms every day, and over 44% of all homicides involve some type of gun violence [1].

Anyone can be affected or threatened by this issue, but in many cases, it tends to impact people of color, women, or other minority groups in far larger numbers. In the U.S., it is also not uncommon for younger children or students to have a larger risk of encountering gun violence at public schools. According to data from Everytown Research, the United States reported an average of over 87 school shootings a year from 2013 to 2021, with an annual average of over 28 dead and nearly 60 injured [2]. Various parents have expressed unease in sending their children to school, an environment that should have been more safe and protected than any other.

Gun violence mainly arises from the easy access to firearms—legal or not—in the United States, which then brings us to the topic of gun control. In recent years, the increase in mass shootings and reports of gun violence in the U.S. have spurred various law enforcement officials and politicians to push for tighter gun control laws.

On June 25th, 2022, President Joe Biden signed into law the first major gun safety legislation that has been passed in decades [3]. This law did not ban any weapons, but included necessary funding for increased school safety and for state crisis intervention programs. Following this trend, many states including California, Delaware, and New York have also passed new laws hoping to reduce gun violence. Some of these laws include regulating

illegally bought ‘ghost guns’, strengthening background check systems to track gun holders, and red flag laws, which allow courts to temporarily seize firearms from anyone believed to be a danger to themselves or others.

According to a study published in January, there is a direct correlation between states with weaker gun laws and higher rates of death caused by gun violence, including homicides, suicides, mass shootings, and accidental killings [4].

1.2 Debates on Gun Control in the U.S.

Some Americans advocate for their right to hold guns, citing the U.S. Constitution as evidence for their free will, while others argue back that even stricter measures regarding gun control saves lives and promotes overall safety. With a debate as large and important as this one, it is all the more important to analyze politicians’ approaches to this in what they say and do, and in doing so, the nation’s overall political state regarding gun control.

1.3 Political Efficacy

Political efficacy, officially defined by Campbell et al. is the “feeling that individual action does have, or can have an impact upon the political process”, or in short, a citizen’s belief in their ability to influence political processes in their state or nation [5]. As such, political efficacy is closely tied with political engagement—understanding of politics, political self-confidence—and the trust in the government’s responsiveness to a citizen’s wants and needs. Higher levels of political efficacy increase citizens’ engagement with political processes, while lower levels of political efficacy lead to alienation from the current political environment [6]. In general, democratic governments are tied to citizens with higher levels of political efficacy, as they have say in the nation’s legislation process [7].

A person’s political efficacy is affected by a large number of variables, but most significantly by the political and social environments in which people live in [8]. This can include the political opinions of those around a person, the political environment

they are in, and the way that politicians speak and present themselves through media. Especially for those that have yet to develop their own politician opinions—in many cases, children—the political opinions of politicians that people around them may support become a large influence in the formation of their own ideas.

Due to this, it is imperative that politicians are conscious and clear in their speech while delivering weighty political statements or ideas, because of the sway and power they hold over both the political atmosphere of the nation and the citizens’ thoughts and responses to it. Recording and analyzing various politicians’ words and opinions has become incredibly important in modern day democratic society, especially regarding a topic as delicate as gun control, which affects the health and safety of all those living in the U.S.

This paper attempts to analyze the speech of various American politicians on the topic of gun control in regards to what opinions they may portray and how they go about presenting these opinions.

2. TF-IDF

TF-IDF, or term frequency-inverse document frequency, is a way to calculate numerical statistics to analyze the relevance of particular ‘terms’ in a document within a collection or corpus. It is widely used across the world with many different uses, including in text analysis, datamining, and is especially useful in machine learning algorithms for natural language processing (NLP).

TF-IDF increases based on the number of times that a term appears in a document, but is also regulated by the number of documents in the group that contain the word. In this way, words that are commonly used in the English language such as the, what, is and other similar words will rank low even though they tend to appear more often compared to other terms. This is helpful considering that in most cases, these words are not relevant to the subjects of the documents. On the other hand, if a more specific term such as ‘fox’ appears a large number of times on a certain document, while seldom appearing in others, the TF-IDF calculation system will rank it very high and deem it a very relevant word.

TF-IDF, while seemingly difficult at first glance, is not too difficult to calculate. TF stands for the ‘term frequency’ of a specific term in a document. While there are many different ways to calculate this frequency, the simplest way is to simply calculate the ratio of the number of times the term appears to

the total amount of words in the document. IDF stands for the ‘inverse document frequency’ of the term across a group or corpus of documents. This will show how common or uncommon the specific term is in the entire corpus. It is calculated by calculating the logarithm of the total number of documents in the group divided by the number of documents that contain the term. In this way, the closer that the IDF is to zero, the more common that the term is within that group of documents. On the flip side, if the term appears much more frequently on a specific document compared to any others, then the IDF value will approach 1.

By multiplying these two values (TF and IDF), the TF-IDF value of any term in a group of documents can be calculated, and can therefore help analyze the frequency and subsequent relevance of the term. The higher the TF-IDF score, the more relevant it is. This is an extremely convenient and easy way to analyze documents quickly for their general ideas and opinions, and can therefore sort and retrieve information much more efficiently. Search engines such as Google use TF-IDF and similar calculations in their algorithms to sort through all of their many documents in order to find and show only the most relevant ones based on the words typed into the engine. Overall, TF-IDF is an essential algorithm to know and understand moving forward.

Example 1. Consider the Documents Below:

Corpus D:

d1 : The fox jumped in the hole.

d2: The fox jumped over the hole.

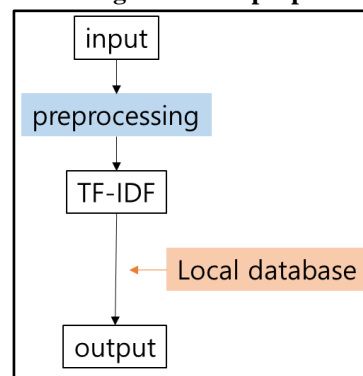
The TF-IDF values would be:

	the	fox	jumped	in	hole	over
TF - IDF	0	0	0	0.05	0	0
TF	2	1	1	1	1	1
IDF	0	0	0	0.301	0	0.301

3. PROPOSED SYSTEM

<figure 1> below shows the overview of the proposed system. The proposed system will consist of an input of data containing twenty speeches of politicians from the United States of America regarding the topic of gun control and gun violence within the U.S.A. Applying the process and the local database with TF-IDF score, the system returns whether the data contains a negative, positive or neutral tone.

Figure 1: A diagram of the proposed system.



3.1 PREPROCESSING

In order to ensure the highest amount of accuracy possible utilizing the TF-IDF algorithm, large amounts of preprocessing was done using the data to ensure accurate TF-IDF values of the documents given. By grouping documents in the corpus into groups of three to five, and running them in turns in the algorithm to check their accuracy, ineffective words were programmed out of the word bag of possible terms. For example, words such as ‘Obama’, which would have been ranked higher in the TF-IDF algorithm due to its moderate rarity and importance in a corpus, but is not relevant to the topic of gun violence itself, would be filtered out through this preprocessing. Other symbols that are frequently used but are not relevant may include the quotation mark symbol, apostrophes, brackets or parentheses, and et cetera. This also entails removing everyday and unnecessary words such as ‘this’, ‘but’, ‘is’, ‘as’, capitalized words that may be ranked higher with no significance such as ‘Obama’ or ‘White House’, and even numbers that have no real meaning, such as ‘twenty-two’ or ‘11’.

Figure 2: Example code showcasing the process of preprocessing the TF-IDF algorithm.

```
def filterP(data):
    formedList = list()
    for d in data:
        d = d.lower()
        d = d.replace(",","")
        d = d.replace("-"," ")
        d = d.replace("--","")
        d = d.replace("—","")
        d = d.replace("0","")
        d = d.replace("1","")
        d = d.replace("2","")
        d = d.replace("3","")
        d = d.replace("/","")
        d = d.replace("president",',')
        d = d.replace("america",',')
        d = d.replace("baltimore","")
```

3.2 TOP-K QUERY

Through this tedious process, the data will be whittled down into a much more concise set of terms ranked in order of significance in regards to the set of speeches in regards to the TF-IDF system called the ‘Top-K’ values. For example, when K is equal to 10, the top 10 values ranked by the TF-IDF system will be arranged from top to bottom in decreasing order, and picked out from the rest of the values.

Table 1: Example Top-K terms (k=3) ranked in descending order of TF-IDF value, sorted into a table.

Top-3 Terms	TF-IDF Value
obsolete	0.001023
gun	0.001011
law	0.000948

As such, the Top-K terms sorted by TF-IDF corresponding to the manually assigned terms in the local database will therefore be assigned values of ‘attitude’ or ‘perspective’ on gun

control, allowing for a better understanding of overall political views on the topic in the eyes of U.S. politicians.

3.3 LOCAL DATABASE

These values will then be compared to the local database. The local database consists of manually picked key words from each document in the corpus, which have each been assigned categories ranging from ‘strongly negative’ to ‘strongly positive’ based on the overall tone of the document containing the word. These categories were assigned after analyzing the tone of each document and matching each key word with the tone of the document it was used in.

Table 2: Example of the assignment of numerical values to Top-K terms based on the local database, sorted into a table.

Key Term	Average Local Database Category	Numerical Value
massacre	strongly negative	-3
law	positive	+2

For instance, the word ‘massacre’ may be assigned a clear -3 from its ‘strongly negative’ value, while the word ‘law’ may be given a +2 value due to its range from ‘neutral’ to ‘strongly positive’.

3.4 RETURNING OUTPUT

The examples below shows how to return the output via this system.

Table 3: Categorization of various Top-K terms in accordance to the local database, sorted into a table.

Strongly Negative	Negative	Weakly Negative	Neutral	Weakly Positive	Positive	Strongly Positive
massacre	N/A	N/A	N/A	N/A	N/A	N/A
N/A	N/A	N/A	law	N/A	law	law

The word ‘massacre’ may be assigned to the ‘strongly negative’ category based on the tone of the document that it was read in. In contrast, the word ‘law’ may be assigned multiple categories ranging from ‘neutral’ to ‘strongly positive’ based on the tones of the relative documents that utilized the word as a key term.

Based on this, each word in the manually picked terms were assigned a corresponding numerical value in the range of -3 to +3 depending on how frequently the words appeared and/or were focused on across documents that may have had negative or positive perspectives on gun control. Negative values will correspond to negative categories in the local database, and the more negative the tone of the document associated with the word, the more negative the value. In this way, the value of -3 would correspond to ‘strongly negative’ terms, -2 would correspond with ‘negative’ terms, -1 with ‘weakly negative’ terms, 0 with ‘neutral’ terms, and so on.

Table 4: Example of overall sortation of Top-K terms (k=3) according to the local database, sorted into a table.

TF-IDF Ranking	Term	Avg DB value	Attitude
----------------	------	--------------	----------

			Value
Top - 1	obsolete	weakly negative	-1
Top - 2	gun	neutral	0
Top - 3	law	positive	2

These values will then be organized into a data set that can be more easily interpreted through use of the local database. This thus makes the output of this proposed system a judgment of the attitudes that the politicians of the United States hold towards the topic of gun control.

4. EVALUATION

The given data was analyzed utilizing the method mentioned above, and 20 tables organizing the 20 documents in the corpus of speeches given by politicians on gun control were created and analyzed as such:

oTable 6: The sorting of Top-K terms (k=10) for document 1 in the corpus regarding the attitude of the speaker on the topic of gun violence, in accordance with the local database, arranged into a table.

Top-K Value	Term	Average Local Database Category	Numerical Value
Top-1	Gun	Neutral	0
Top-2	People	Neutral	0
Top-3	Weapon	Negative	-2
Top-4	Ban	Strongly Negative	-3
Top-5	Children	Strongly Negative	-3
Top-6	Assault	Strongly Negative	-3
Top-7	Protect	Negative	-2
Top-8	School	Strongly Negative	-3
Top-9	Duty	Negative	-2
Top-10	Health	Negative	-2

*Total Numerical Value: -20

**Overall Attitude: 20% neutral, 40% negative, 40% strongly negative

Results conclude a variety of perspectives on gun control and gun violence across multiple different politicians, with an overall negative outlook on gun violence. Numerical values for

attitudes range from +5 to -25, with a much more heavy leaning towards negative values, and thus toward a negative outlook on gun violence across the U.S..

Table 7: Overall attitudes of politicians in the U.S. regarding gun control in terms of numerical value in accordance with the local database of Top-K terms (k=10) of all documents in the corpus.

Document Number in Corpus	Numerical Attitude Value Based on Local Database
Document-1	-20
Document-2	-25
Document-3	-13
Document-4	-16
Document-5	-10
Document-6	-15
Document-7	-17
Document-8	-17
Document-9	+5
Document-10	-20
Document-11	-19
Document-12	0
Document-13	-4
Document-14	-6
Document-15	-14
Document-16	+1
Document-17	-19
Document-18	-11
Document-19	-17
Document-20	-13

From this data, it can be concluded that a majority of politicians in the U.S. hold negative opinions toward gun violence, and hold concern about how the current state of gun laws in the country will affect citizens in the future. Points of most concern, and the

most heavily mentioned terms analyzed by the TF-IDF algorithm seem to be ‘children’, ‘school shootings’, and ‘mental health’ in documents with more negative attitudes. Documents with more positive attitudes emphasize ‘freedom’, ‘individual

rights', and seem to argue that gun laws 'infringe' on the right to bear arms written in the constitution.

As such, due to the majority of attitudes in the corpus being negative and thus, supportive of increased gun control in the U.S., it can be concluded that a majority of politicians in the U.S. currently hope to adjust gun laws to enforce more control upon the average population in terms of access to firearms in hopes to decrease the amount of deaths and mass shootings, especially in schools and in places where children may come to harm. Considering the common mention of the words 'mental' and 'health', many other politicians seem to be considering increasing support and encouraging the growth of mental health facilities in the aspect of prevention of possible future gun violence.

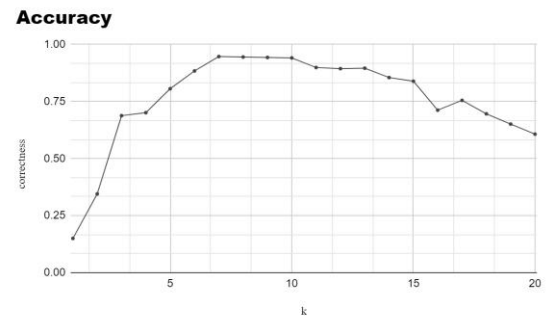
The minority of politicians concerned with the loss of individual rights to bear arms place importance on the terms 'constitution' and 'self-defense'. It appears as though a number of politicians who are gun owners themselves are concerned with the loss of personal liberty and rights to self-defense, arguing that the right to defend oneself in case of emergency, and therefore the right to bear arms, is an inherent right established within the federal constitution. However, these views tend to stray towards a more neutral tone as seen in the data, with only a +5 being the most strongly positive attitude in comparison to politicians whose attitudes on gun violence remain more negative, whose tones range from -4 all the way to -25, with a majority of them exceeding -10.

The rest of the documentation evaluations can be found:

https://docs.google.com/document/d/1HCzT5VVQJ9W_Hyo_8UOi7Ft0RTG8Ex7JtHSqsD1vvnw/edit?usp=sharing

Furthermore, I have also conducted the accuracy evaluation. First, to propose a proper k value I observed the accuracy and the k value.

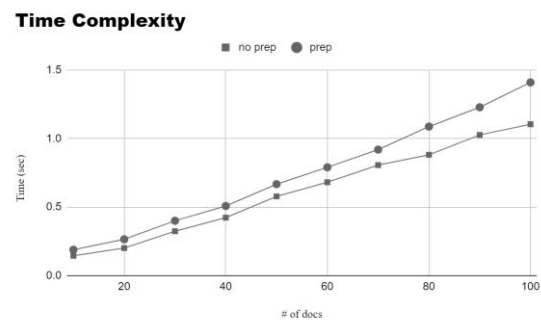
Figure 3: Graph showcasing the relative accuracy of the TF-IDF algorithm in relation to k-values.



By observing the <figure 3> above, it can be seen that starting from k=7, the accuracy stays at a constant peak for a consistent amount of k values. At this point, the maximum accuracy of 0.947% can be achieved, proving that utilizing k values of k=7 to k=10 will provide the highest accuracy of the TF-IDF program. However, in order to maximize efficiency of time, future investigations should utilize the value of k=7 in order to achieve the best results possible.

Second, to ensure this system is fit to the real world, the time complexity evaluation was conducted with and without preprocessing.

Figure 4: Graph showcasing the time complexity of the TF-IDF algorithm with and without preprocessing and with increasing numbers of documents.



According to <figure 4> above, it can be observed that documents that have undergone preprocessing will always take more time to process in comparison to documents without preprocessing due to the higher complexity and longer length of code. Furthermore, the graph shows that as the number of documents run in the algorithm increases, the time taken increases as well in a general increasing slope.

5. CONCLUSIONS AND FUTURE

6. WORKS

In conclusion, the topics of gun violence and gun control remain widely debated and a major problem in the United States of America. Various politicians can be seen to hold a variety of opinions and perspectives on the topic, all of which are eminent in their tone of voice as they discuss the state of the U.S. concerning gun control. Through this study, it could be seen that a majority of politicians hold negative views on the current political landscape of the U.S. in terms of gun violence, and are concerned about the safety of citizens and children in particular without any further changes in firearm regulation. Many opinions on how gun laws and how the right to bear arms should be regulated have been expressed, including focusing more political attention on mental health facilities by prioritizing prevention of gun violence, or restricting gun ownership and tracking of firearms more heavily through consistent background checks and more thorough gun control laws.

By utilizing the TF-IDF algorithm in order to explore and analyze Top-K values of these politicians' speeches, an efficient and systematic way of organizing data regarding political views on gun violence and the opinions of specific politicians in the U.S. regarding this topic were able to be found. The usage of the TF-IDF algorithm in this method proved to be not only efficient, but also much easier to understand, able to analyze 20 medium-length speeches in just a few minutes by viewing only the key terms and phrases emphasized within the documents.

More exploration in this field of programming should be considered in future studies in order to analyze the accuracy and clarity of the TF-IDF algorithm across a variety of genres and topics.

7. REFERENCES

[1] "Gun Violence." *Amnesty International*, Amnesty International, www.amnesty.org/en/what-we-do/arms-control/gun-violence/.

[2] "School Shootings by Country 2022." *World Population Review*, World Population Review, worldpopulationreview.com/country-rankings/school-shootings-by-country.

[3] Tucker, Emma. "These are the gun control laws passed in 2022 so far." *Cable News Network*, Cable News Network, 30 July 2022, edition.cnn.com/2022/07/30/us/gun-control-laws-2022/index.html.

[4] "Everytown Gun Law Rankings." *Everytown*, Everytown for Gun Safety Support Fund, 2022, everytownresearch.org/rankings/.

[5] Schafer, Mark, et al. "Democratic leaders and the democratic peace: The operational codes of Tony Blair and Bill Clinton." *International Studies Quarterly*, vol. 40, no. 3, 2006, pp. 561-583. <https://doi.org/10.1111/j.1468-2478.2006.00414.x>

[6] Stoker, Gerry. *Why Politics Matters*. 2nd ed., Palgrave, 2017.

[7] Morrell, M.E. "Deliberation, Democratic Decision-Making and Internal Political Efficacy." *Political Behavior*, vol. 27, 2005, pp. 49-69. <https://doi.org/10.1007/s11109-005-3076-7>

[8] Anderson, Christopher J.. "Economic voting and political context: a comparative perspective." *Electoral Studies*, vol. 19, ed. 2-3, 2000, pp. 151-170. [https://doi.org/10.1016/S0261-3794\(99\)00045-1](https://doi.org/10.1016/S0261-3794(99)00045-1)