"Public Sentiment on Nashville School Shooting on Reddit": Sentiment Analysis Using Empath and Topic Modeling Using Latent Dirichlet Allocation (LDA) (Word Count: 2848)

Name: Srikeerthana Reddy Bandi
Course: Master's in Data Science
Semester: Spring'23

1. Introduction:

On March 27, 2023, a mass school shooting occurred at The Covenant School, a Presbyterian Church in America parochial elementary school in the Green Hills neighborhood of Nashville. The shooter, Aiden Hale, a former student at the school and a transgender man, killed six people, including three nine-year-old children and three adults, before being shot and killed by the police. This tragic event has sent shockwaves through the Nashville community, sparking discussions about gun violence, school safety, and mental health. The incident has also brought to the forefront the complex intersection of transgender rights and mental health with gun control policies.

As the community continues to process this devastating event, it is crucial to understand the public's sentiments and opinions surrounding it. The Nashville school shooting has once again highlighted the urgent need for effective measures to prevent gun violence and ensure the safety of children in schools. Social media platforms such as Reddit provide a wealth of information about people's opinions and emotions related to the shooting. According to a study by Kalesan et al. (2018), school shootings have become a significant public health issue in the United States, with an increase in the frequency and severity of incidents in recent years. Additionally, Wang et al. (2016) analyzed Twitter sentiments after the 2012 Sandy Hook Elementary School shooting and identified varying public responses from each state, based on gun ownership levels. Their study provides a useful framework for analyzing sentiments expressed on social media platforms in response to school shootings.

Understanding how the public views this tragedy is essential for policymakers and community leaders in their efforts to address these issues and prevent similar incidents in the future. As such, I plan to conduct an analysis of public sentiment and perform topic modeling to uncover key themes and insights that reflect people's thoughts and feelings about this event. By doing so, we can gain valuable perspectives on this tragedy and work towards creating a safer and more secure environment for our schools and communities.

2. Research Question:

1. What is the public sentiment on Reddit regarding the recent school shooting in Nashville?
2. What are the primary topics surrounding the Nashville school shooting being discussed on Reddit?

3. Method:

3.1 Data:

The dataset for this research was collected using PRAW, a Python package designed for accessing the Reddit API. The data was collected using the PRAW library, by filtering subreddits related to

"Nashville school shooting", "covenant school shooting", and "mass shootings". The choice of subreddits was made after careful consideration of their relevance to the research question, and our aim was to include a diverse range of perspectives on the topic. Thus, I chose to analyze comments from four different posts, each offering a unique viewpoint.

The four posts were collected manually by taking the below facts into consideration for my analysis:

- A post discussing the last message from the shooter.
- A post discussing the idea of limiting news coverage on mass shootings especially school shootings.
- A post discussing the possibility that the shooter targeted the school due to being bullied for being transgender.
- A post discussing Senator Josh Hawley's comments blaming the shooting on Christian hate.

These posts were chosen because they represent a variety of perspectives and ideas related to the shooting, including discussions of the shooter's motivations, potential ways to prevent future shootings, and political reactions to the event.

The PRAW library in python was used to extract comments from a specific Reddit post and save them to a CSV file. The script first writes the header row to the CSV file, then extracts the post's title, author, score, creation date, and number of comments and writes them as the first row of data in the file. Next, the script iterates through all the comments in the post, recursively handling any nested replies. For each comment, it extracts the author's username (or indicates if the comment was deleted), score, creation date, and the comment text itself, and writes them as a row in the CSV file. I used the date attribute to record the timestamp for each comment, to analyze the comments within a specific timeframe. I have executed this process for all four posts and merged their respective comment data into a single CSV file for further analysis.

Title	Author	Score	Date	Comment
CMV: We should stop doing widespread news coverage on mass shootings (school shootings to be specific).	*3xtheredcomet*	*45*	*4/1/23 22:38*	*One youtuber I used to watch a lot, Philip Defranco, covers the news including mass shootings, but his policy is to never mention the killer by name nor put their picture on screen. I feel like that's a decent compromise, what do you think?*
The last message from Audrey Hale, the Covenant School shooter	*DonZinger*	*438*	*3/28/23 5:05*	*Why take people out with you? Makes no sense. There has to be more to this than just "I wanted to die".*
Is it so wrong to theorize that the Nashville gunman probably shot up the Christian school because he was bullied for being transgender?	*Hopeful-Base-2769*	*0*	*3/31/23 12:46*	*If not all mentally I'll people want to shoot up schools; the same is true that not all people who own guns don't want to randomly kill innocent people. GUNS ARE FOR PROTECTION, NOT AGGRESSION. Most transgender people aren't even Christian & shouldn't care what Christians believe. The Bible says that same sex orientation is a sin & that's okay. Take that issue up with God through prayer; not killing innocent children. Pastors marry men & women through Biblical principles & standards. Redefining the "male/female" terms of marriage does NOT make a Pastor/Priest obligated to marry anyone that defies the Biblical definition of marriage. And that's okay; just go to someone else to get married. Leave Christians alone; it's not our fault that's what the Bible says about it.*
Hawley blames Nashville School shooting on Christian hate	*victrasuva*	*268*	*3/28/23 13:57*	*But when it happens at a public school 'it's no time for politics'. If only there was something we could do to stop these mass shootings.... Thoughts and prayers obviously aren't working.*

Figure 3.1.1: Illustrates the collection of comments through Python libraries from four different posts.

The timeframe for the dataset was from the day of the shooting March 27th, 2023 to present. This timeframe was chosen to capture comments related to the immediate aftermath of the shooting and any ongoing discussions related to the event.

Overall, our dataset consists of the title, author, score, date, and comment of posts that contained the specified keywords and fell within the specified time frame. I have gathered a dataset consisting of 1631 comments from four different posts using PRAW.

3.2 Preprocessing:

In order to prepare the data for sentiment analysis and topic modeling, preprocessing on the text data was performed. The preprocessing steps included removing duplicates, converting all text to lower case, removing punctuations and numbers, eliminating URLs, tokenizing the text into individual words, and removing stop words. Additionally, excluded comments that were deleted or removed to ensure that only relevant data is considered for analysis.

To perform preprocessing, a python script was executed. The script uses the csv module to read and write CSV files. It opens the 'total_comments.csv' file in read mode, creates a CSV reader object, and skips the header row. The script uses the pandas, re, and nltk libraries to remove URLs, punctuations, and numbers, convert text to lowercase, tokenize text into individual words, remove stop words, and join the tokens back into a string. It then removes deleted or removed comments from a CSV file containing comments data, applies the preprocess_text function to the 'comment' column, and drops duplicate comments. The preprocessed data is saved to a new CSV file.

The preprocessing steps are necessary to clean and standardize the text data before it can be used for sentiment analysis and topic modeling. Removing duplicates ensures that the same comment does not affect the analysis multiple times. Converting all text to lower case standardizes the text and ensures that words are not treated differently because of capitalization. Removing punctuations and numbers helps to focus on the actual text and removes distractions that do not add value to the analysis. Eliminating URLs removes any external links that may not be relevant to the analysis. Tokenizing the text into individual words helps to break down the comments into meaningful units for analysis. Finally, removing stop words removes commonly used words like "the" and "and" that do not provide any specific meaning and may skew the results. Excluding deleted or removed comments ensures that only relevant data is considered for analysis and avoids any irrelevant or misleading information. This preprocessing step was necessary to prepare the data for further analysis and to gain valuable insights into the sentiments expressed in the posts. Post, preprocessing, I have a total of 1546 posts to analyze.

3.3 Analysis:
Sentiment Analysis:

To examine the public sentiment towards the recent Covenant school shooting which took place in Nashville, I conducted sentiment analysis using the Empath tool. Empath was chosen because it is a widely used and reliable tool for sentiment analysis, and it provides a comprehensive list of categories to analyze emotions and sentiment. The tool uses a large list of keywords and categories to identify emotions and sentiment in text data, making it a valuable tool for analyzing public opinions and attitudes towards different topics. With the help of Empath, insights into the general

sentiment and emotions expressed by the public in response to the Covenant school shooting were gained.

Pandas library was used to read our preprocessed file, which contained 1546 data points (comments) for sentiment analysis. First, using Empath, calculated the positive, negative, neutral, for each comment. Empath counts the number of words associated with each category and calculates a score for that emotion. In the case of positive, negative, and neutral emotions, Empath counts the number of words associated with positive, negative, and neutral categories, respectively. For example, words such as "happy" and "love" are associated with positive emotions, while words such as "hate" and "angry" are associated with negative emotions.

If a text contains more positive-related words, the positive score will be higher. If a text contains more negative-related words, the negative score will be higher. If a text contains an equal number of positive and negative-related words, the neutral score will be higher. one more column was added to our data frame to classify the sentiments as positive, negative, or neutral. To accomplish this, following categorization rule is used.

Positive score	Negative score	Neutral score	Categorized as
Higher than both negative and neutral scores	-	-	*positive*
-	*Higher than both positive and neutral scores*	-	*negative*
-	-	*Higher than both positive and negative scores*	*neutral*

Table 3.3.1: Illustrates the categorization of scores.

These are some sample comments along with their corresponding sentiment, analyzed using Empath.

S.No	Comment	Positive score	Negative score	Neutral score	Emotion
1	*one biggest sparks movement black rights us murder year old emit till mother insisted public funeral service open casket saying let people see boy tens thousands attended funeral images bloated mutilated body published black oriented magazines newspapers rallying popular black support white sympathy across u uvalde shooting pediatrician attended victims described small bodies pulverized decapitated image seeing images bodies little children like tv instead coverage outside people crying maybe extreme bottom line think way lessen school shootings hide problem anything horrendous effects need substantially amplified action taken prevent*	*0*	*2*	*0*	*negative*
2	*us st amendment protects one cherished freedoms freedom speech banning news coverage shooting would limitation freedom ban also would take away public right know dangers facing country democracies work proprerly public know issues need solved public know school shootings problem vote anti gun candidate ban news mass shootings allows great argument ban news reporting subjects name prevention public safety could ban news police shootings leads riots could ban news train accidents prevent panic*	*3*	*0*	*0*	*positive*
3	*someone hide guns*	*0*	*0*	*0*	*neutral*
4	*large group psychologists submitted petition say media coverage infamy motivation attacks laws passed prevent coverage freedom speech hold point speech causing others harmed danger*	*1*	*0*	*0*	*positive*
5	*please share voting record mental health legislation thanks*	*0*	*0*	*0*	*neutral*

Figure 3.3.1: Illustrates the scores and sentiment of the posts using Empath.

Considering the recent Covenant school shooting that has deeply impacted the United States, it is essential to gain a comprehensive understanding of the emotions expressed by the public regarding the event, in addition to analyzing comments as positive, negative, or neutral. To achieve this, I utilized Empath, a tool capable of categorizing emotions and providing valuable insights into how individuals are feeling.

A Python instance of the Empath library was created, read comments about the Covenant school shooting from a CSV file, and defined a list of appropriate emotions, including sadness, depression, dread, anxiety, horror, fear, love, hope, trust, surprise, boredom, pain, aggression, and disgust. I have added columns to the Data Frame for each relevant emotion. The next step was to loop through each remark and use Empath to analyze the emotions that were present in each one. The updated Data Frame contained the matching scores. After that, a column was added for each comment's dominant emotion, which was identified as the emotion with the greatest score. If there was no dominant emotion, the comment was classified as neutral.

comment	sadness	depression	dread	anxiety	horror	fear	love	hope	trust	surprise	boredom	pain	aggression	disgust	ominant_emotio
also sure exactly think hand thousands families con	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*2*	*0*	*aggression*
think people committing deplorable acts want answ	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*disgust*
hawley embarrassment missouri understand live mi	*0*	*0*	*0*	*0*	*0*	*2*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*fear*
people shootings want die want die ensure spread	*4*	*0*	*0*	*0*	*0*	*2*	*2*	*0*	*1*	*0*	*0*	*4*	*1*	*3*	*sadness*
government ban medium expression used citizenry	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	*0*	*surprise*
speaking school shooters discussion agree gun viole	*1*	*0*	*0*	*0*	*0*	*0*	*2*	*0*	*1*	*0*	*0*	*2*	*1*	*0*	*love*
agree right problem violence like sells news paper	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*	*0*	*0*	*0*	*1*	*0*	*trust*
derail topic son man executed promoted rank admir	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*neutral*

Figure 3.3.2: Illustrates the sample comments, scores, and dominant emotion of the comments using Empath.

Topic Modeling:

Topic modeling is an essential technique in the field of machine learning and natural language processing. Its purpose is to identify latent topics within large text documents. With the growth of social media platforms, researchers are increasingly interested in extracting meaningful insights from user-generated content. Posts and comments on these platforms are often unstructured and short, making topic modeling a valuable tool for discovering underlying themes and ideas within them. According to Hamed Jelodar et al. (2017), there are various methods for topic modeling, among which Latent Dirichlet allocation (LDA) is one of the most popular methods in this field.

I conducted topic modeling to gain a better understanding of the different perspectives surrounding the Nashville school shooting incident. I collected comments from four different Reddit posts discussing the incident and used the LDA algorithm, a popular machine learning technique that identifies the primary topics within a text corpus by analyzing the word frequency. As the text had already been preprocessed for sentiment analysis, I used the same preprocessed text for LDA analysis but also removed stop words again for greater accuracy.

First, the preprocessed text was converted into a bag-of-words corpus format. Then, the LDA model was trained on this corpus with a specified number of topics. Finally, by examining the top 10 words for each identified topic, I gained insight into the main themes and concerns expressed in the comments related to the Nashville school shootings. This process allowed me to delve deeper into the underlying themes and concerns expressed in the comments.

4. Results:

A total of 1546 data points were analyzed. Following are the graphs which visualize the percentage distribution of scores using Empath.

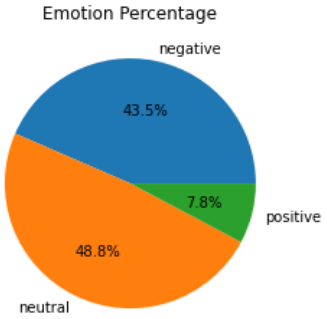

Figure 4.1: Pie Chart representation of sentiment analysis using Empath

The pie chart shows the results of sentiment analysis on a set of data using the Empath. Of these, 7.8% are classified as positive sentiment, 43.5% as negative sentiment, and 48.8% as neutral sentiment for Empath.

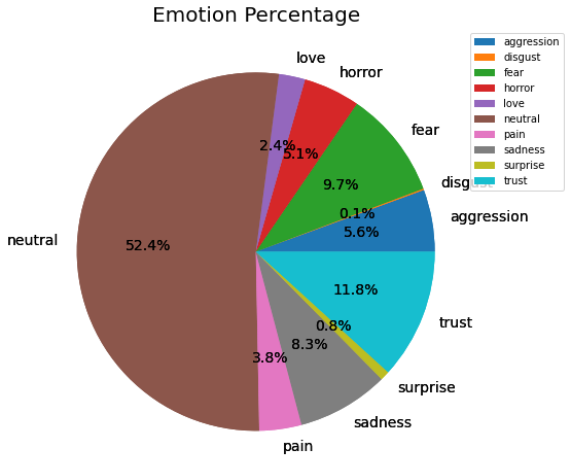

Figure 4.2: Pie Chart representation of distribution of emotions

According to the pie chart, the emotions most associated with the Nashville school shooting are trust, fear, and sadness, with aggression and pain also playing a significant role. It's clear that negative emotions are overshadowing the positive emotions. The relatively low score for love suggests that this emotion is not a prominent theme in discussions about the shooting, and there are also very few mentions of disgust, surprise, or boredom. Overall, the data indicates that the Nashville school shooting has evoked strong emotions such as fear, sadness, and trust in people, which is likely a reflection of the gravity of the situation and the profound impact it has had on the community. It's worth noting that the high percentage of neutral emotions (52.4%) may be due to the inclusion of comments that don't fit neatly into the categories provided.

Statistic	Value
Mean	168.38
Median	94.00
Standard deviation	234.79
1st percentile	4.00
99th percentile	1076.64

Table 4.1: Key statistics of the distribution of comment length and frequency

The mean comment length is 168.38, indicating that on average, the comments have 168.38 words. The median comment length is 94.00, meaning that half of the comments have fewer than 94 words and half have more than 94 words. The standard deviation of the comment length distribution is 234.79, indicating the degree of variability or dispersion of the comment length values. The 1st percentile value of 4.00 means that 1% of the comments have fewer than 4 words, while the 99th percentile value of 1076.64 indicates that 99% of the comments have fewer than 1076 words.

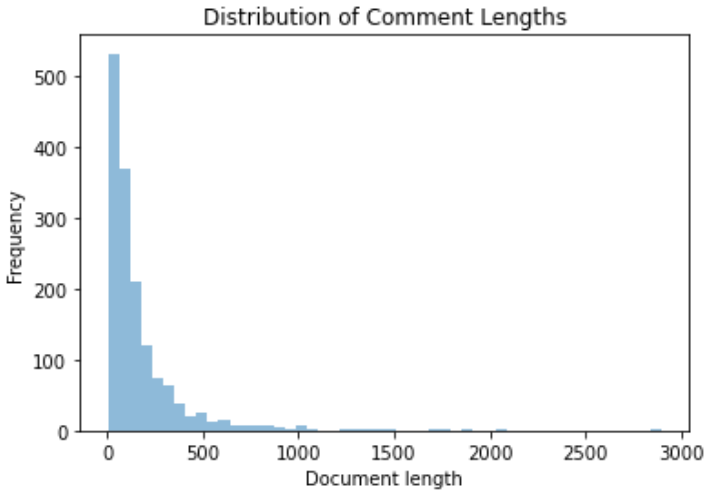

Figure 4.3: Distribution of Comment lengths

RQ2: Topic modeling

Figure shows the result of a word cloud image of all the comments created using the WordCloud module in python.

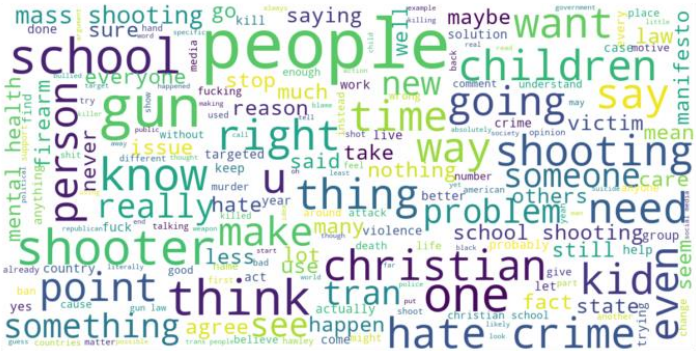

Figure 4.4: Word Cloud illustrating summary of the most frequently occurring words.

The word cloud provides a visual representation of the most frequently used words in our data. As we can see, it covers all the key themes and concepts related to the Nashville school shooting, such as ‘shooting’, ‘school’, ‘gun,’ ‘Christian’, ‘hate’, ‘shooter’, ‘mental health’, and ‘trans’. These keywords help to provide an overall understanding of the context and the issues surrounding the topic. By highlighting the most used words, the word cloud helps to emphasize the central ideas and themes that are relevant to the Nashville school shooting.

```
Topic: 0
Words: 0.018*"people" + 0.016*"trans" + 0.006*"way" + 0.005*"say" + 0.005*"mental" + 0.005*"one" + 0.005*"want" + 0.0
05*"person" + 0.004*"reason" + 0.004*"guns"

Topic: 1
Words: 0.018*"gun" + 0.013*"guns" + 0.009*"us" + 0.009*"people" + 0.009*"shootings" + 0.006*"news" + 0.006*"like" +
0.006*"school" + 0.006*"mass" + 0.006*"countries"

Topic: 2
Words: 0.016*"school" + 0.015*"people" + 0.012*"hate" + 0.009*"like" + 0.008*"shooter" + 0.008*"crime" + 0.008*"thin
k" + 0.008*"gun" + 0.007*"christian" + 0.006*"one"

Topic: 3
Words: 0.015*"people" + 0.007*"like" + 0.007*"one" + 0.006*"guns" + 0.006*"school" + 0.005*"mental" + 0.005*"problem"
+ 0.005*"get" + 0.004*"shootings" + 0.004*"health"

Topic: 4
Words: 0.014*"people" + 0.007*"school" + 0.007*"think" + 0.007*"shooter" + 0.006*"want" + 0.006*"like" + 0.006*"know"
+ 0.006*"news" + 0.005*"shooting" + 0.005*"manifesto"
```

Figure 4.5: The topics generated using LDA model.

Through trial and error, I settled on using 5 number of topics for my topic modeling. After experimenting with different numbers of topics ranging from 3 to 10, it seemed that the best results were obtained with 5 topics. The resulting frequencies and topics are presented below, generated using LDA.

The LDA analysis of the Nashville school shooting data produced four topics: transgender rights and mental health, gun violence and mass shootings, hate crimes and the Christian community, and a general topic on people's thoughts and opinions related to mental health and gun control policies.

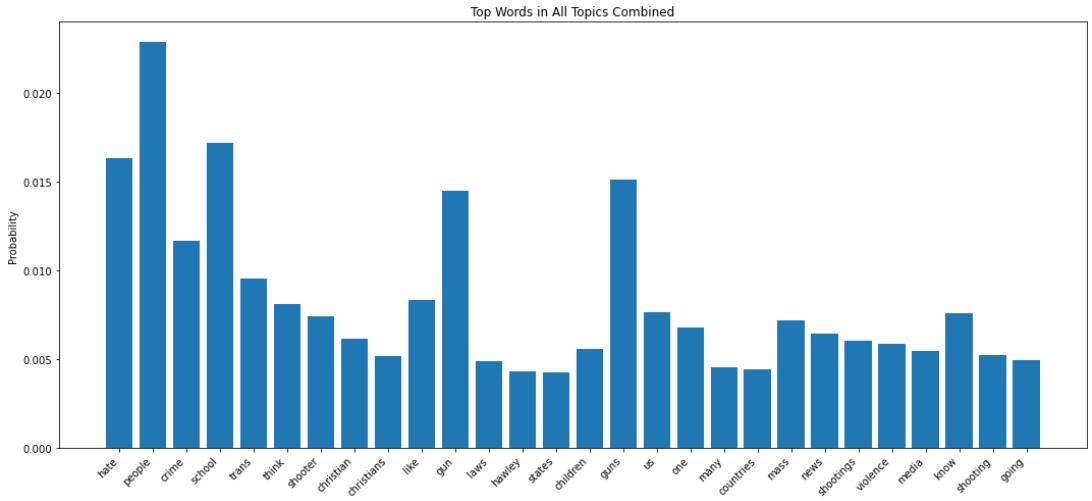

Figure 4.6: A bar plot of the top words in all topics combined in a trained LDA model.

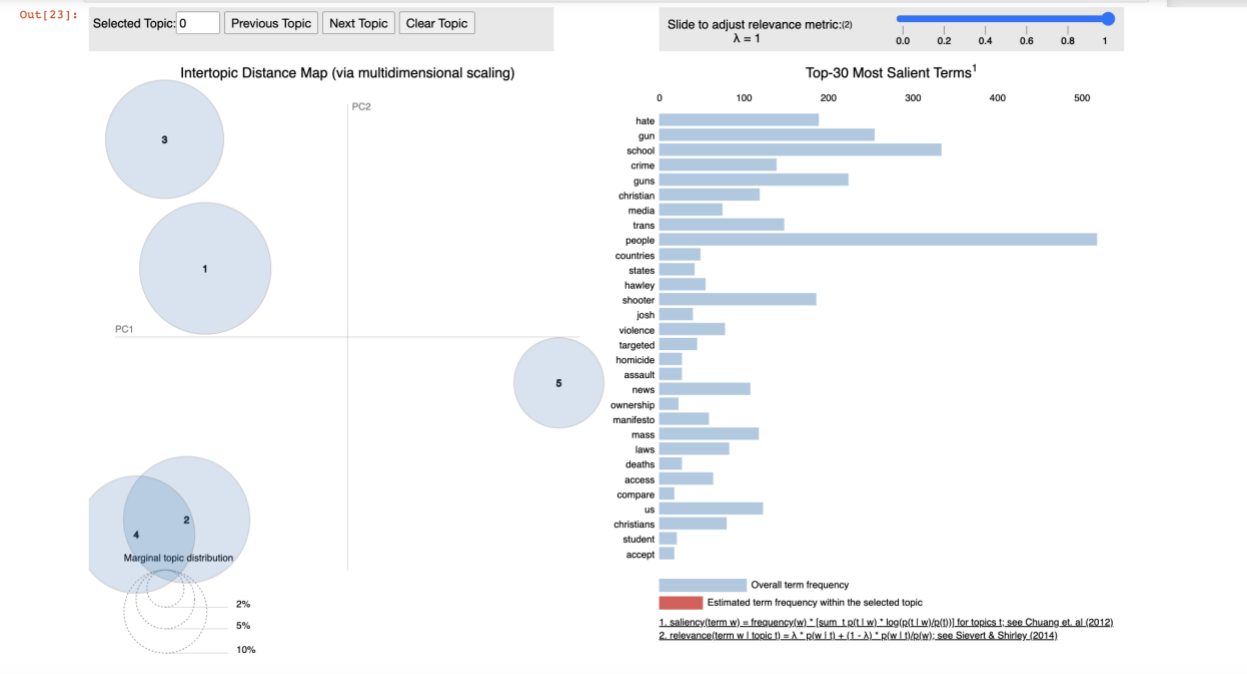

Figure 4.7: Visualization of Top 4 topics using LDA

pyLDAvis is a helpful tool for visualizing and interpreting the results of a topic model. It provides an interactive visualization that allows users to explore the underlying topics and their relationships.

5. Conclusion:

In conclusion, the sentiment analysis and LDA topic modeling of data related to the Nashville school shooting have provided valuable insights into the emotional responses and themes associated with this tragic event. The dominant negative emotions of fear, sadness, and trust reflect the severity of the incident and its impact on the community. The LDA topics identified important discussions about various social and political issues related to gun violence, mental health, hate crimes, and transgender rights. These insights can provide valuable information for policymakers and organizations to better understand the public's response to tragic events such as school shootings and work towards implementing effective solutions.

Overall, the analysis of online discussions related to the Nashville school shooting highlights the importance of understanding the broader social and political contexts in which such incidents occur. It also emphasizes the need for greater attention and action to address issues related to gun violence, mental health, hate crimes, and transgender rights. By gaining a deeper understanding of the emotions and themes associated with tragic events like the Nashville school shooting, we can work towards creating a safer and more equitable society.

6. Limitations:

Limitations of the Data: The data used in this study is limited to 1000 data points and may not be able to represent the broader population's sentiments and opinions.

Limitations of Sentiment Analysis: The sentiment analysis using Empath only provides a broad understanding of the emotions expressed in the discussions. It does not provide insight into the context and intensity of those emotions.

Limitations of LDA: Additionally, the LDA topics are generated based on the frequency of words and may not capture the full meaning and intent of the discussions.

7. References:

1) Kalesan, B., Lagast, K., Villarreal, M., Pino, E., Fagan, J., & Galea, S. (2018). School shootings during 2013-2015 in the USA. Injury Prevention, 24(4), 261-267. https://doi.org/10.1136/injuryprev-2016-042162

2) Wang, N., Varghese, B., & Donnelly, P. D. (2016). A machine learning analysis of Twitter sentiment to the Sandy Hook shootings. https://doi.org/10.1109/eScience.2016.7870913

3) Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2017). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. https://doi.org/10.48550/arXiv.1711.04305

4) Hernandez, J. (2023, March 28). What we know about the deadly shooting at a Nashville elementary school. NPR. https://www.npr.org/2023/03/28/1166482479/nashville-school-shooting-covenant-what-we-know

5) Prabhakaran, S. (2018, December 4). Topic modeling visualization – How to present the results of LDA models? Retrieved from https://www.machinelearningplus.com/nlp/topic-modeling-visualization-how-to-present-results-lda-models/