

Full name and semester: Srikeerthana Reddy Bandi - Spring'23

Title: “Analyzing Public Perception Towards Nashville School Shooting on Reddit through EMPATH and Topic Modeling” (Word Count: 636)

Research Question:

1. How do individuals express their sentiments towards the recent school shooting in Nashville in the comments section of Reddit posts?
2. What are the key themes and topics that are related to Nashville school shooting?

Background: The recent Nashville school shooting on March 27th, 2023, has once again brought the issue of gun violence in schools to the forefront. 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. The studies indicate the importance of analyzing the discussion on social media platforms like Reddit about the Nashville school shooting to gain valuable insights into public attitudes and perceptions regarding the issue of gun violence in schools.

Data: To gather the necessary data for analysis, I will utilize PRAW, a Python library that enables me to interact with Reddit's API. PRAW facilitates access to the Reddit API using authentication Client and Secret ID's. Initially, I will conduct a search by writing a python code for posts related to the recent Nashville school shooting using keywords such as Nashville, school shooting, and gun violence. Based on the results retrieved from the code, I will select a post that has most number of comments using sort functionality in Reddit. Once I locate the post, I will use PRAW to extract a minimum of 1000 comments from it. To ensure accuracy in sentiment analysis, I will preprocess the collected data by removing URLs, HTML tags, and special characters. The scraped data will include relevant information such as the username, comment, time, and date, and upvotes. Since the shooting occurred on March 27th, 2023, I will limit my search to posts published on or after that date.

Method: For this study, I will tokenize the text using NLTK's word_tokenize function to convert the text into a list of words or tokens, which is essential for preparing the text data for analysis. Sentiment analysis will be used to classify the tone of the collected data. The data will be analyzed using, Empath tool, a Python library that allows for the identification and extraction of emotional categories from text data. EMPATH analyzes text and assigns weights to various emotions to generate an overall polarity score that ranges from -1 to +1. Categories with high scores contribute to a positive sentiment, while low scores contribute to a negative sentiment. This approach goes beyond simple positive or negative sentiment analysis and can detect a wide range of emotions such as anger, fear, and sadness. EMPATH sentiment analysis will be applied to each comment, and a new column containing the sentiment score for each comment will be added to the data frame. This is the approach I am trying to perform.

In addition, I am planning to analyze the key topics and themes that are related to school shootings in Nashville using Latent Dirichlet Allocation (LDA), a topic modelling technique. By applying LDA to the corpus of text, we can identify the most likely topics that are present in the data. It may help us to analyze the textual data related to Nashville School shootings, where we can extract the most probable topics that are present in the corpus of text related to school shootings. Some of these topics may involve debates surrounding the government's responsibility, gun violence, the role of parents and teachers in monitoring children, and other related subjects. The focus of this study will be on exploring the potential of using empath sentiment analysis and topic modeling to understand how individuals express their emotions and sentiments towards the recent school shooting in Nashville.

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>