Sentiment Analysis of Israel-Hamas Conflict in Gaza

Determining overall public sentiment via Reddit regarding the ongoing conflict in Gaza between Israel and Hamas

December 8, 2023

Prepared By:

Devanshi Shah
Quantitative Methods in the Social Sciences

Introduction

The objective of this sentiment analysis is to draw conclusions about the general public opinion regarding the ongoing conflict between Israel and Hamas (a Palestinian militant group), which has recently escalated following the October 7th attack on Israel by Hamas. Following this attack, Israel (via the Israeli Defense Forces) launched a counter-offensive in Gaza, which has come under fire for violating numerous international human rights laws that amount to war crimes, such as the blockage of electricity and water into Gaza and the bombing of refugee camps. Since October 7th, escalated fighting between Israel and Hamas has been ongoing, with any hopes of a permanent or long-term ceasefire dismal. Due to the scale of and history behind this conflict, it has attracted international attention from the media, governments, and civilians. Civilians specifically have been very active on social media platforms, posting their views about the conflict. One such platform, Reddit, contains much of this sentiment and is the medium utilized to determine public opinion about the Israel-Hamas conflict and humanitarian issues in Gaza.

Methods: Text Collection and Data Pre-processing

The post selected from Reddit is titled "Israel orders more Gazans to flee, bombs areas where it sends them." The post has a total of 2,301 comments (excluding duplicate comments). To clean the data before sentiment analysis was conducted, the desired variables for each comment were extracted and placed into a data frame. This included the actual text of the comment, the number of "ups" and "downs" per comment, and the date created. The dates were then converted into a human-readable format (YYYY-MM-DD), the duplicate comments were dropped, and the final dataset was saved as a CSV file. Then, the CSV file was loaded back into the notebook, and the unnecessary columns were dropped from the data frame. After that, the necessary libraries and collection of stopwords were imported so that a function could be created that passes through all the comments and removes any unnecessary words and symbols to make sentiment analysis more efficient. The comments were then tokenized (split into individual words), and any words in a longer form than the root were stemmed using the PorterStemmer() function. Upon completion of all the aforementioned steps, the data becomes ready for sentiment analysis.

Sentiment Analysis

First, a word cloud (Fig. 1) was generated using the cleaned comments data. The most commonly used words included "Israel," "Hamas," "Gaza," "Palestinians," "civilians," and "Jews." The full word cloud can be viewed in the Appendix. Furthermore, a bar chart (Fig. 2) of the most common words broken down by frequency to understand the magnitude of word use can also be found in the Appendix.

Next, the TextBlob library was imported to determine the polarity and subjectivity of the comments. The numeric polarity values for each comment (-1 indicates high negativity, 1 indicates high positivity) were classified into 3 categories: Negative, Neutral, and Positive, in which neutral comments are closer to a score of 0. The same categorization method was applied to the subjectivity score of the comments, in which values greater than 0.5 are considered subjective, while values at 0.5 and below are considered objective. The frequency tables (Figures 3 and 4) in the Appendix display the counts for each of the three types of sentiments and types of subjectivity. Considering the former, there were more positive comments (1001) than negative (694) and 606 neutral comments. The polarity scores were then displayed in a histogram (Fig. 5), which indicates that most of the sentiment lies more in the slightly positive region, with a mostly normally distributed shape. Looking at subjectivity, 1,505 comments were objective, while 796 were subjective. The same histogram was created for comment subjectivity scores (Fig. 6), and while most of the graph appears

normally distributed around 0.5, there is a high frequency of objective statements with scores between 0 and 0.1.

Next, descriptive statistics were run for polarity and subjectivity scores (the table, Fig. 7 is in the Appendix). From the statistics, it is clear that the comments are largely neutral on average (around 0.03), while for subjectivity, the comments are light to moderately subjective, with a mean score of around 0.39. The standard deviations for both polarity and subjectivity were quite similar as well (about 0.245 and 0.28, respectively). The minimum and maximum values for each measurement reached the absolutes: -1 and 1 for polarity, and 0 and 1 for subjectivity. Finally, a crosstab was created to determine the variance of subjectivity by polarity across all categories of each variable (Fig. 8). Most of the objective statements were neutral or positive (570 and 586 comments, respectively), while most of the subjective statements were either negative (345) or positive (415), with very few being neutral (36). On the other hand, negative sentiment was mostly evenly split between objective and subjective (349 and 345, respectively), neutral sentiment was overwhelmingly objective (570 vs. 36), and positive comments were slightly more objective overall (586 vs. 415).

Implications

It is not surprising that the two most commonly used words are “Israel” and “Hamas,” as these are the two main actors involved in the conflict. What is notable is that words that are related to Palestinian civilians and Gaza (not Hamas), though in the top ten most frequently used words, are used far less than the top two to three words (in the 200-300 range vs. the 900-1000 range, respectively). This indicates that Reddit users could likely be more interested in posting opinions related to the conflict itself, such as logistics, tactics, military assaults, etc., rather than issues more centered around the humanitarian crisis civilians are facing in Gaza. It is also quite surprising that the polarity scores are somewhat normally distributed, and that the mean polarity score is approximately 0.03, given the divisive nature of the conflict. It was expected that there would be more negative sentiment, on average, given the destruction and violence that has ensued in Gaza since October 7th. Perhaps this can be reconciled with the notion that users are more likely to post fact-based updates about the military aspect of the conflict rather than providing opinions about the state of civilians in Palestine.

The subjectivity results of the analysis were also unexpected, again, given the controversy behind the conflict. The mean was only about 0.39, indicating that the average comment was subjective, but not more than moderately so. There was an expectation that generally, comments would be more subjective (i.e. a mean score of 0.5 or above), but, interestingly, that the sentiment is not strongly subjective. This could reflect a desire from users to post accurate information to avoid the spread of misinformation, as well as a weariness of attracting negative attention for highly subjective comments from those who may disagree. Another intriguing finding was within the histogram for subjectivity scores- the data does not appear to be normally distributed, unlike the histogram for polarity, as there is a high frequency of objective statements with scores between 0 and 0.1, while the rest of the scores are more concentrated around 0.2 to 0.7. This outlier group likely brought the average down.

From the crosstab, it seems reasonable that a vast majority of the subjective comments were either negative or positive rather than neutral, as neutral comments reflect a more objective tone overall, while subjective comments likely contain stronger-toned words to encompass users’ opinions. Furthermore, it is also reasonable that the neutral sentiments were overwhelmingly objective, given this rationale. Overall, it appears that public sentiment on Reddit about the Israel-Hamas conflict is largely neutral and slightly subjective, but this sample size is very narrow, and thus sentiment data on this conflict from other sources should be extracted and aggregated with the Reddit data for a more accurate analysis of public opinion.

Appendix

Figure 1: Word Cloud

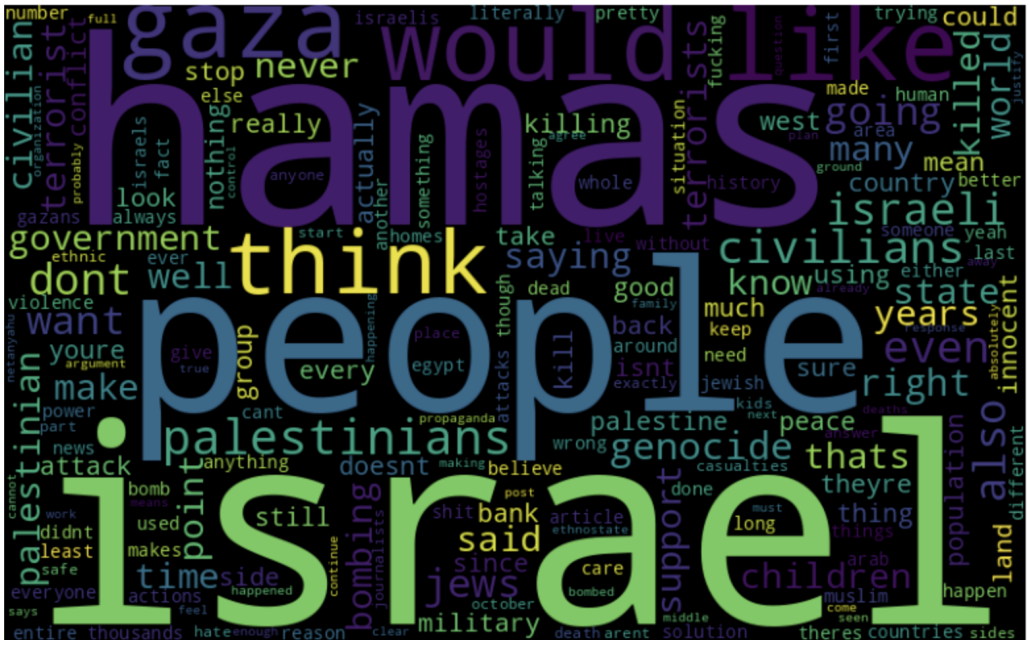

Figure 2: Bar Chart of Frequently Used Words

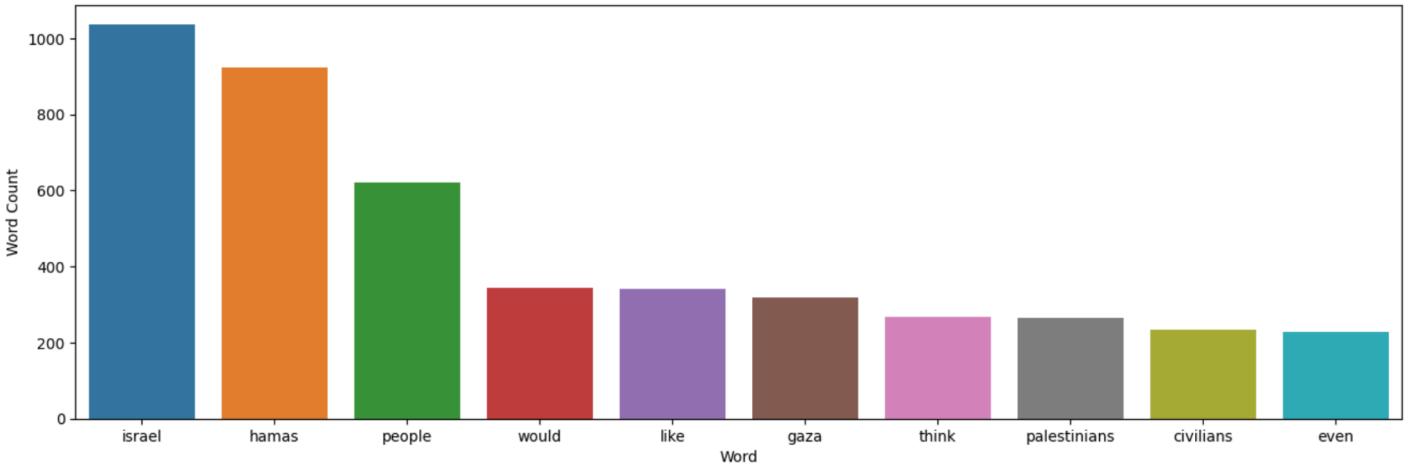

Figure 3: Frequency tables for polarity categories

Sentiment Type	Count
Positive	1001
Neutral	694
Negative	606

Figure 4: Frequency tables for subjectivity categories

Subjectivity	**Count**
Subjective	1505
Objective	796

Figure 5: Polarity Histogram

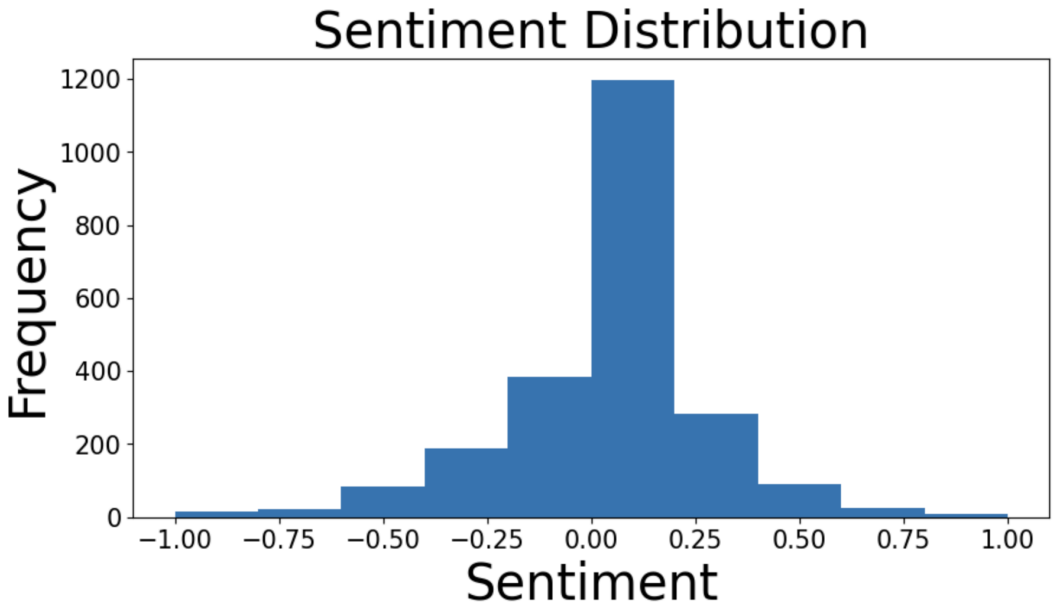

Figure 6: Subjectivity Histogram

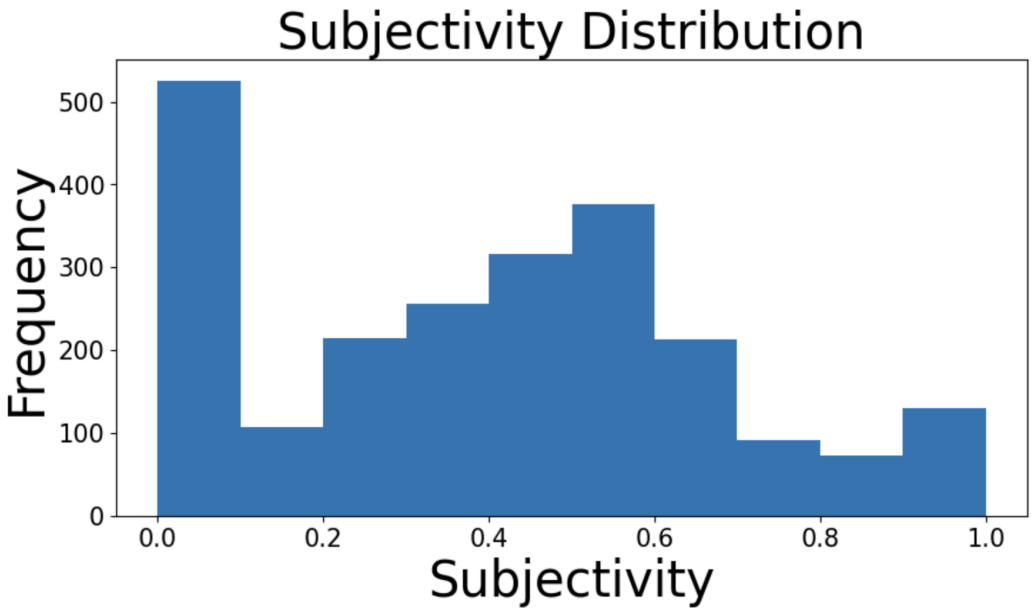

Figure 7: Descriptive Statistics on Polarity and Subjectivity

	polarity	subjectivity
count	2301.000000	2301.000000
mean	0.026997	0.388200
std	0.245113	0.282299
min	-1.000000	0.000000
25%	-0.050000	0.125000
50%	0.000000	0.400000
75%	0.150000	0.577778
max	1.000000	1.000000

Figure 8: Crosstab for Polarity and Subjectivity Categories

subjectivity_binary sentiment	Objective	Subjective
Negative Sentiment	349	345
Neutral Sentiment	570	36
Positive Sentiment	586	415