Sentiment Analysis from Twitter for March 2022 of Ukraine-Russia Conflict

Author: Dayvonn Jones, Nilay Vinchhi

Abstract

Twitter is one of the many organizations and social media platforms that have been exploited by and misused by organizers and supporters on either side of the conflict. People hope to use the platform to express beliefs and ideas that are not verified or peer-reviewed. Textual data were collected to observe negative and positive sentiments in the collection of tweets during March 2022 about the conflict in Ukraine. It can come from a popular outlet or an unverified user. The data was cleaned and then categorized to extract sentiments from the textual data. The sentiments and ideas expressed were visualized with a dashboard.

Introduction

Social media is a powerful tool for expressing the freedom of speech. Especially in countries with no freedom outreach, either such media are banned or monitored. We are in a unique situation with the Ukraine-Russia conflict; it is our modern internet of things driven world's first war. The war intensified when Russia declared the eastern part of Ukraine Donbas, whose conflict has been evident since Donetsk and Luhansk were recognized as independent by Russia. (Platonova, 2021) Social media has documented a significant amount of activity in the Russian-Ukrainian conflict, which began in February 2022. Multiple news outlets across the world have reported that entities within Russia have exploited misinformation in journalism and cybersecurity vulnerabilities. (Sabatovych, 2019) One overwhelming theme is that it associates organizations with origins in Russia that have organized and targeted organizations and groups in other countries to spread misinformation and retrieve sensitive information. Twitter is one of the many organizations and social media platforms that have been exploited by and misused by pro-Russian supporters. If there is a positive sentiment tweeted about the war in Ukraine, it will most likely be from an account with little engagement and obscure history; a bot account. (Polyzos, 2022)

There was a war before in 2014 when Russia seized part of the country in response to a civil overthrow. In the war, Russia wants the current Ukrainian regime to stop persecuting Pro-Russian groups based in the country. Russia invaded the country in the hopes of demilitarizing the current reigning government and imposing or placing a government that shares an interest with Russia (Kirby 2022). The first few battles of the conflict occurred in 2014 and continued in 2022. We have lost thousands of lives in the current situation, and the way has mental, physical, and economic effects on the residents and the world. We are going to analyze the sentiments

Hypothesis

The citizens living in the warzone of Ukraine have emotions that need to be expressed to understand the issue of modern war. If the citizens in an affected area have access to utilities (like the internet), they can share a more accurate description of their sentiment and updates on the war.

Method and Current Sentiment Analytics

The data was analyzed solely with the software R Studio. Data on engagement statistics were collected from the dataset that contained the number of retweet counts and the number of replies and was sorted from verified and unverified accounts. Data cleaning was performed to extract the user has provided location from the Twitter API rtweet. Data cleaning includes Filtering the dataset into March 2022; The ggplot2 library was used to express the engagement tweets by user verification. A bubble chart was created to express the volume of tweets with the sentiment of the war based upon the geolocation, using Kepler.gl. The sentiment.ai package was used to score the text from the tweets and arranged by negative and positive comments on the ongoing conflict.

For the engagement statistics part, we collected the statistics from the dataset with the number of retweet counts and the number of replies, and they were sorted from verified and unverified accounts. We have the variable that signifies if the account was Twitter by a verified account by Twitter or an account that was not recognized and verified by Twitter

Data was collected using a Twitter API retweet, and it was determined that the full text of a tweet was essential to analyzing the sentiment. There were several data queries to collect a total of 1,400 tweets. The tags that we used to collect the tweets were "#ukraine," "#russia," "#russian," "#stopputin," "#stoprussia," "#ukrainerussiawar," "#standwithukraine," "#ukrainearmy," "#donetsk," followed by the name of cities with 250,000+ population as per census 2021, Ukraine.

Queries were specified by organizing the source for the geolocation in Ukraine. Twitter often provided a User's geolocation if the user consented to share the location. The collected tweets were merged into one file, and then data cleaning was performed to provide pertinent information such as engagement, full text, time of tweet, geolocation, and user profile. Additionally, we prepared a network analysis for hashtags, representing thicker lines with more number relationships between the tags, the largest circle represents the most frequent tag, and the smaller represents the opposite. We can observe from the hashtag analysis that there are two islands of networks; the first one is about Ukraine and its war from a supportive perspective as a whole, and the second cloud is about Donetsk; the reason that these tags are separate might be because the eastern region is as it is the first where it was attacked, and where Russia declared the war.

We used four different approaches to capture the emotions for our Sentiment Analysis. The packages we used are sentiment.ai, syuzhet, and tidytext. In tidytext, we used two different dictionaries, i.e., bing lexicon and lourghan lexicon. Meanwhile, most sentiment analysis models use a dictionary or directly pick up the words; they do not understand the context of the sentence and how we use a combination of words to mean something different. Even though sentiment.ai is a bit different from other dictionaries, it has a dictionary, but it has a deep learning model, which analyzes the context of the text and joins them into a sentence, and then alot a phrase rate sentiment score to that. Additionally, the library also supports multilingual dictionaries, so we can also get sentiment scores and phrases for foreign language tweets, including Ukrainian and Russian.

The world of natural language processing has advanced it, but the most standard approach of NLP is to confirm the sentences with a library and pick up the sentiment phrase out of it, and that is what we have done for syuzhet tidytext. It is a different debate to discuss which dictionary fits which better; it is mostly based on the data you are inputting; not one dictionary can be perfect for all. (Giachanou & Crestani, 2016) Here, we selected the get_nrc_sentiment() to get a data frame represents a sentence from the sentiment phrase that we got from sentiment.ai, a column will choose if it is positive or negative and will filter them into these classes of emotion, "anger", "anticipation", "disgust", "fear", "joy", "sadness", "surprise", "trust", "negative", "positive.". Hence, we got emotions from the sentiment phrase; the reasoning for selecting this methodology was a limitation of dictionary-based libraries, which are mostly inclined towards sentiment analysis in the English language. The libraries in tidytext represent a similar fashion; they have two different dictionaries and different row values to represent emotions. Finally, we used the phrases for each sentiment analysis to get emotions. Hence, our dataset will not have much contradiction, as it is all based on the data generated in sentiment.ai, even though it supports the limitation of not cross-correlating our results.

A word cloud was created using ggwordcloud to illustrate the sentiment phrases generated by library words that are scored using the deep learning library in sentiment.ai, The frequency of those phrases is used in the Twitter data. We also prepared a dashboard using Flexdashboard, which uses Plotly's interactive visualizations to describe our dataset in three sections. First, the Sentiment Analysis plotted on the X-axis is time, Y axis as sentiment scores, and the tweets are plotted using the Viridis package's color scheme. Further, we used crosstalk to create a shiny interactive app with a slider, selector, and type to search button. Secondly, we represent the spatial location of the tweet's user-provided geolocation and tweet counts about the place. In the third section, we describe the primary statistical distribution of the tweets, which are visualized using the Plotly library.

Discussion: Limitations and Opportunities

Data containing languages that were a part of the data and not in the English language were removed in dictionary libraries, but in the sentiment.ai textual information is translated for sentiment analysis. Sentiment Analysis uses different dictionaries; meanwhile, each dictionary has a set of words that captures its sentiment score; we can get a gist of the sentiment phrase. The geolocation provided is user-provided geolocation, which depicts there can be a false entry. The geolocation was matched by the top 250,000+ population cities' names in Ukraine, 2021. Hence, one limitation can be that if there are any tweets about less populous cities, they will not be in our dataset, for example, snake island.

Conclusion and Policy Implications

Natural language processing has advanced to read and interpret human emotions successfully; there are several ways a machine can interpret human language, like facial expressions analysis, heart rate sensor, eye tracking sensor, physical trackers, and almost every combination of wearable technology. However, the purpose of this is to measure people's emotions, and government can use it to analyze how people are feeling in a quantifiable measure (scores). This tool analyzes the texts, here in terms of tweet texts; the government can also use the comments they receive from any survey campaign and analyze the emotions of residents to a specific program. Additionally, we also generated sentiment phrases, which describe explicitly what the tweet's sentiment is describing. Apart from using it just as a tool, sentiment analysis is an advancing field to understand how the human brain works and how it works; This discussion leads to the application of the human-trained dataset to create a human alike artificial intelligence model or (HAI) (Xing et al., 2022). Twitter is an excellent medium to share, as well as we can take the benefit from its hashtag feature. (Asiaee T. et al., 2012) There is also evidence of depression detection using social media from textual analytics. (Varghese Babu et al., 2022) Sentiment Analysis can also help get or estimate an event; if we see more negative sentiment tweets, it depicts there is tension about the given topic/ hashtag. By this methodology, we can forecast based on temporal patterns of a series of sentiments for a given tag or set of tags.

References

- Asiaee T., A., Tepper, M., Banerjee, A., & Sapiro, G. (2012). If you are happy and you know it... tweet. ACM International Conference Proceeding Series, 1602–1606. https://doi.org/10.1145/2396761.2398481
- Drieger, P. (2013). Semantic Network Analysis as a Method for Visual Text Analytics. Procedia - Social and Behavioral Sciences, 79, 4–17. https://doi.org/10.1016/j.sbspro.2013.05.053
- Giachanou, A., & Crestani, F. (2016). Like It or Not. ACM Computing Surveys (CSUR), 49(2). https://doi.org/10.1145/2938640
- Kirby, P. (2022, April 29). Why has Russia invaded Ukraine and what does Putin want? BBC News. Retrieved May 6, 2022, from https://www.bbc.com/news/world-europe-56720589
- Platonova, D. (2021). The Donbas Conflict in Ukraine : Elites, Protest and Partition. The Donbas Conflict in Ukraine. https://doi.org/10.4324/9781003213710
- Polyzos, E. (2022). Escalating Tension and the War in Ukraine: Evidence Using Impulse Response Functions on Economic Indicators and Twitter Sentiment. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.4058364
- Redondo, T., & Sandoval, A. M. (2016). Text Analytics: the convergence of Big Data and Artificial Intelligence. International Journal of Interactive Multimedia and Artificial Intelligence, 3(Special Issue on Big Data and AI). https://www.ijimai.org/journal/bibcite/reference/2540
- Sabatovych, I. (2019). Do social media create revolutions? Using Twitter sentiment analysis for predicting the Maidan Revolution in Ukraine. Global Media and Communication, 15(3), 275–283. https://doi.org/10.1177/1742766519872780
- Varghese Babu, N., Grace, · E, & Kanaga, M. (2022). Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review. 3, 74. https://doi.org/10.1007/s42979-021-00958-1
- Xing, F. Z., Cambria, E., & Welsch, R. E. (2017). Natural language based financial forecasting: a survey. Artificial Intelligence Review, 50(1), 49–73. https://doi.org/10.1007/s10462-017-9588-9Analytics. IEEE Access, 8, 142173–142190. https://doi.org/10.1109/access.2020.30139

Appendix

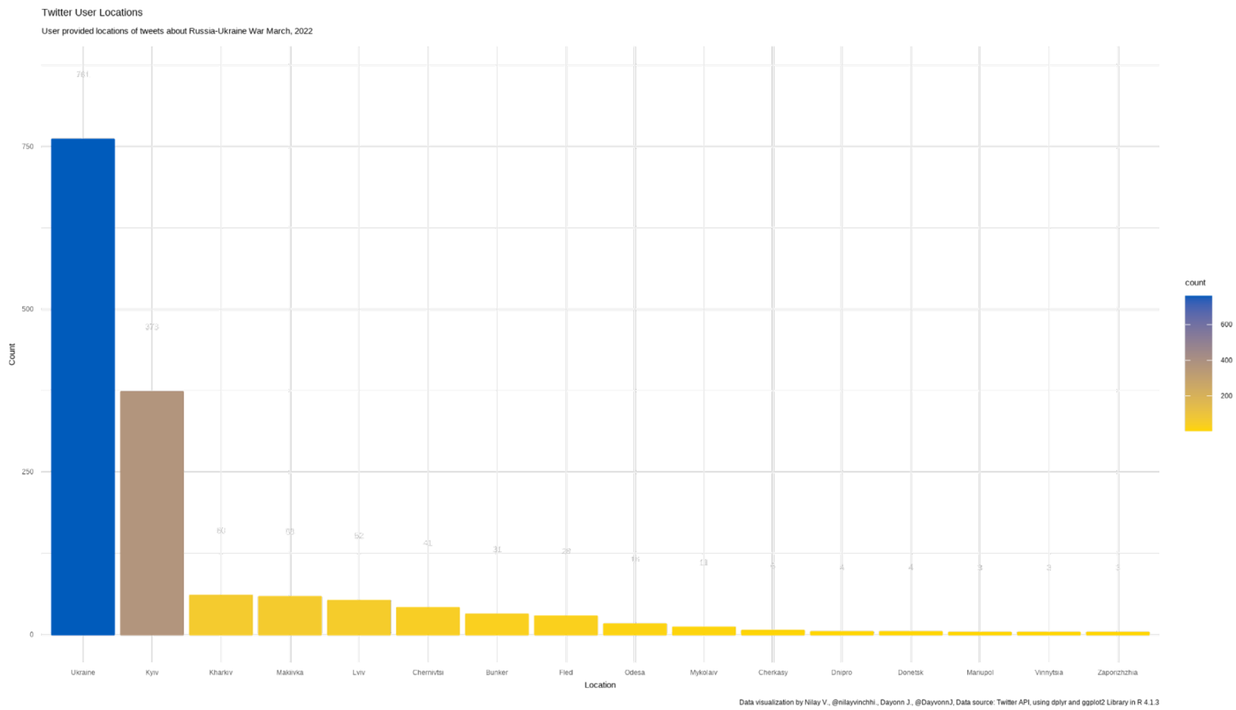

Figure 1: User-Provided Location Bar Chart by Count for tweets in Ukraine about the Ukraine-Russia war 2022

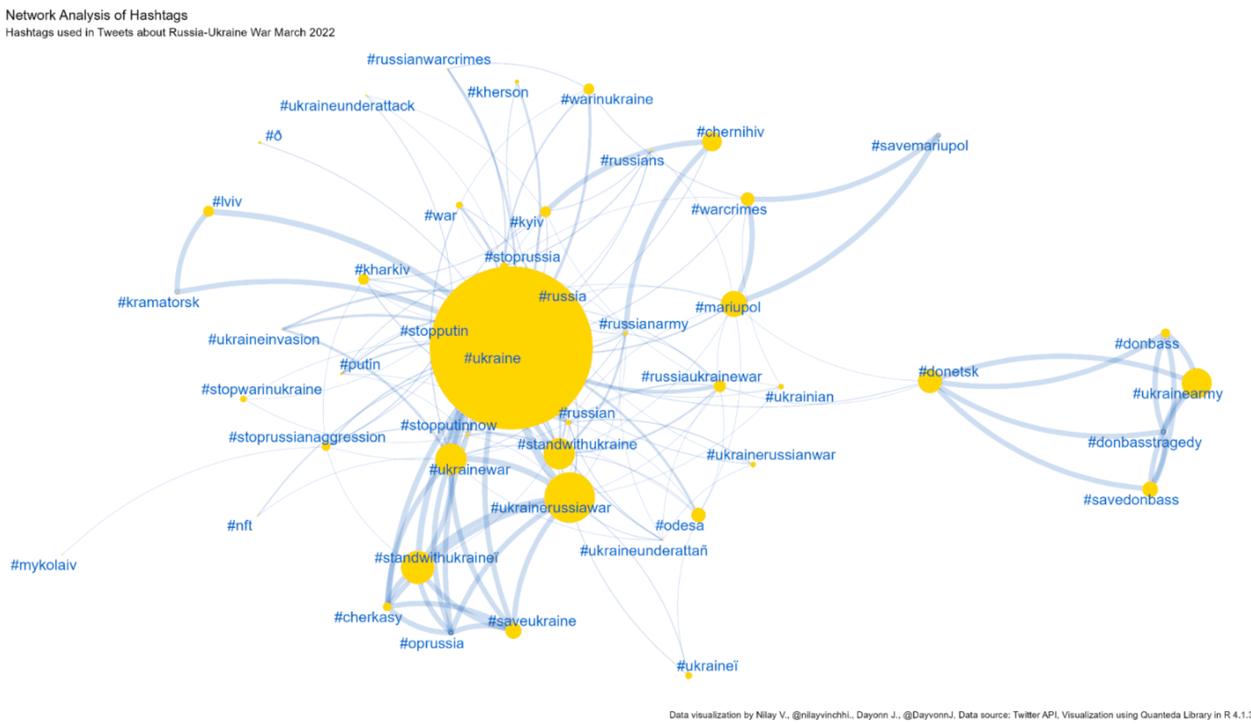

Figure 2: Hashtag Network Analysis for tweets in Ukraine about the Ukraine-Russia war 2022

Figure 3: Wordcloud by positive-negative Sentiment Class for Sentiment Phrases for tweets in Ukraine about the Ukraine-Russia war 2022

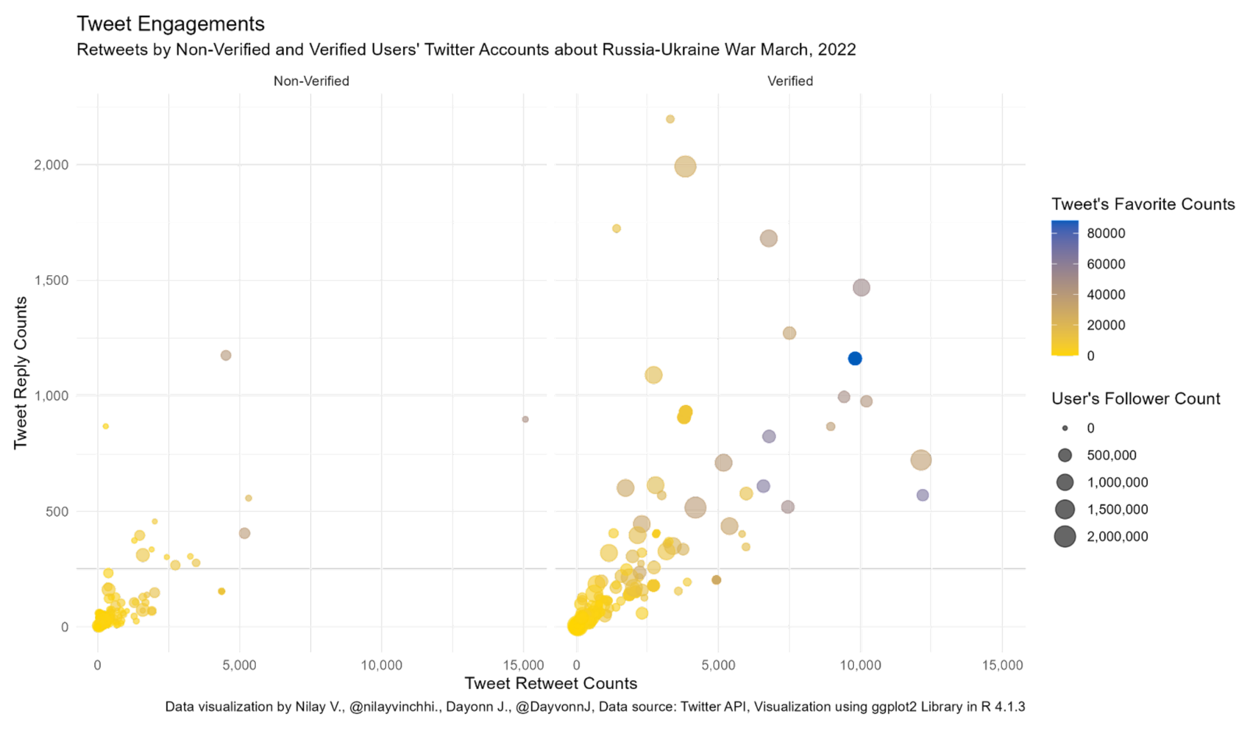

Figure 4: Tweet Engagements: Retweets by Replies, b Verified and unverified Twitter account for tweets in Ukraine about the Ukraine-Russia war 2022

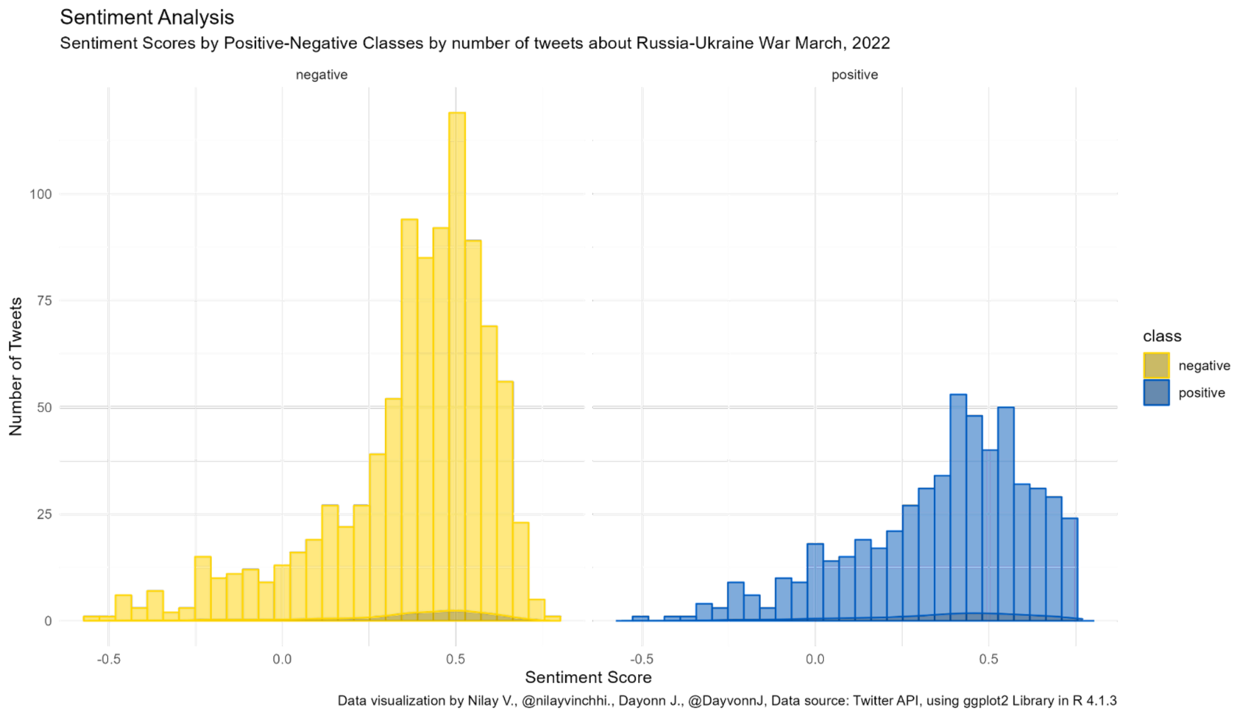

Figure 5: Sentiment Class by Count for tweets in Ukraine about the Ukraine-Russia war 2022

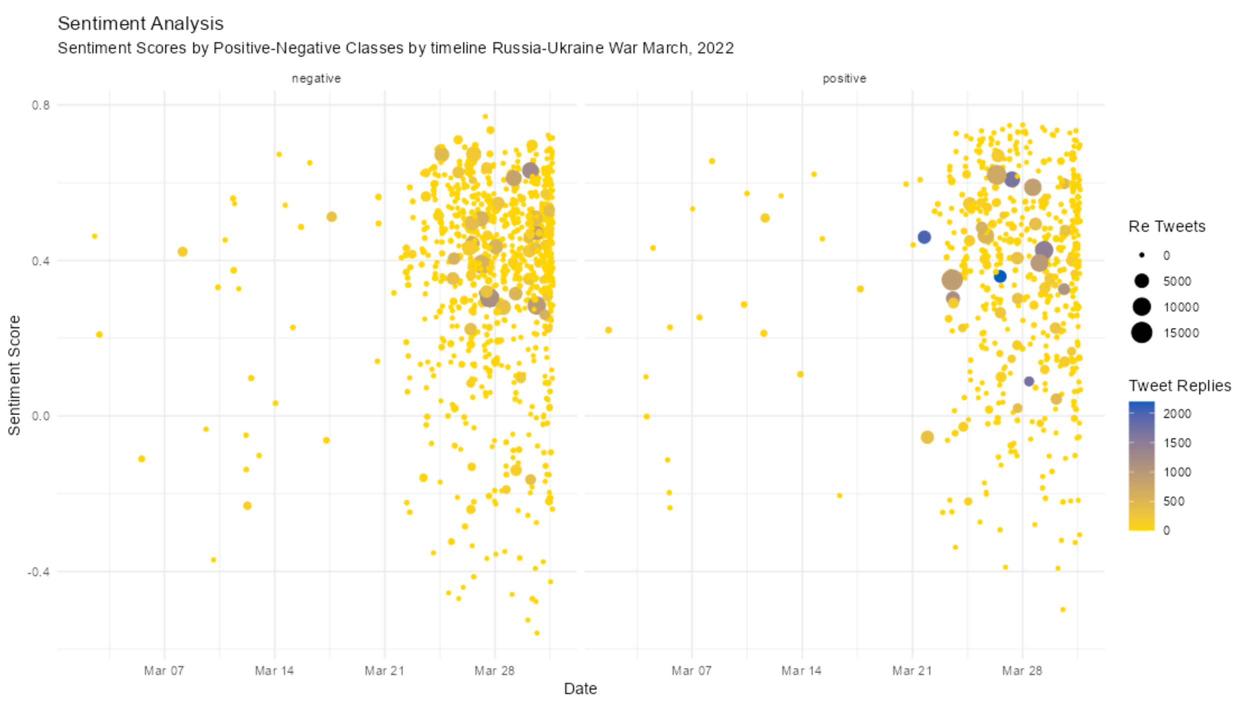

Figure 6: Sentiment Score by date for negative-positive tweets in Ukraine about the Ukraine-Russia war 2022

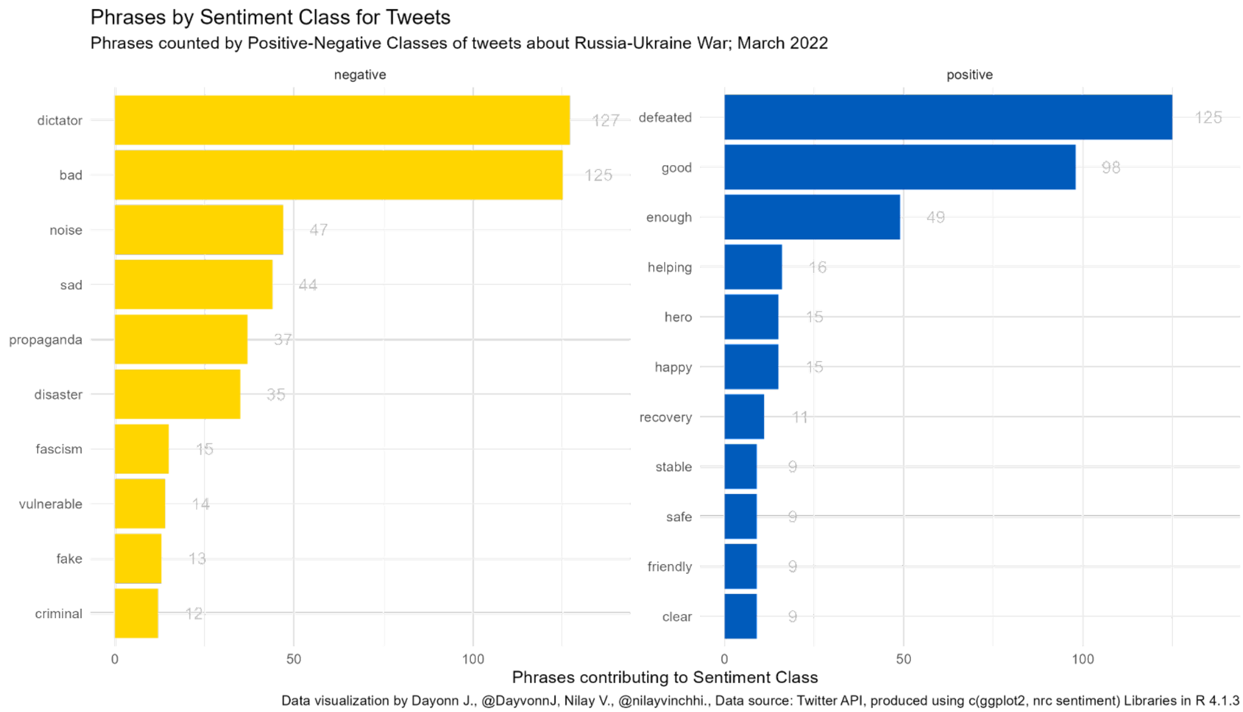

Figure 7: Emotions count using syuzhet library for tweets in Ukraine about the Ukraine-Russia war 2022

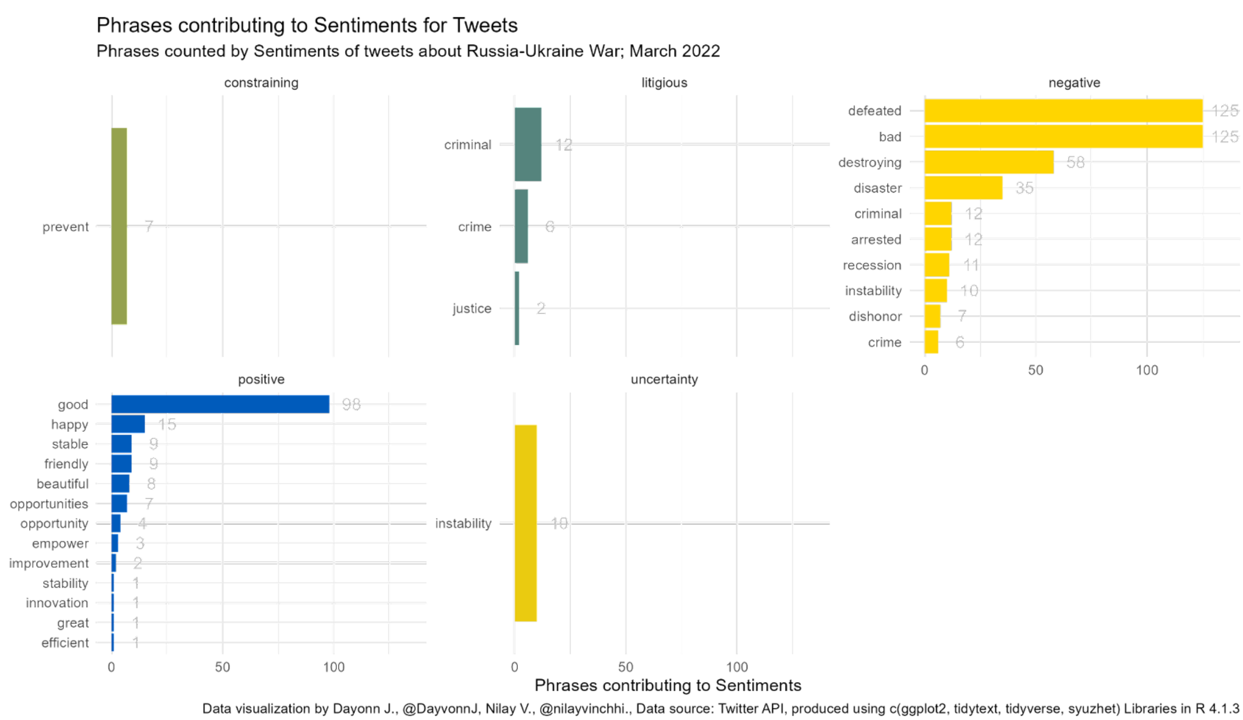

Figure 8: Sentiment Analysis using Lourgan Lexicon by Count for tweets in Ukraine about the Ukraine-Russia war 2022

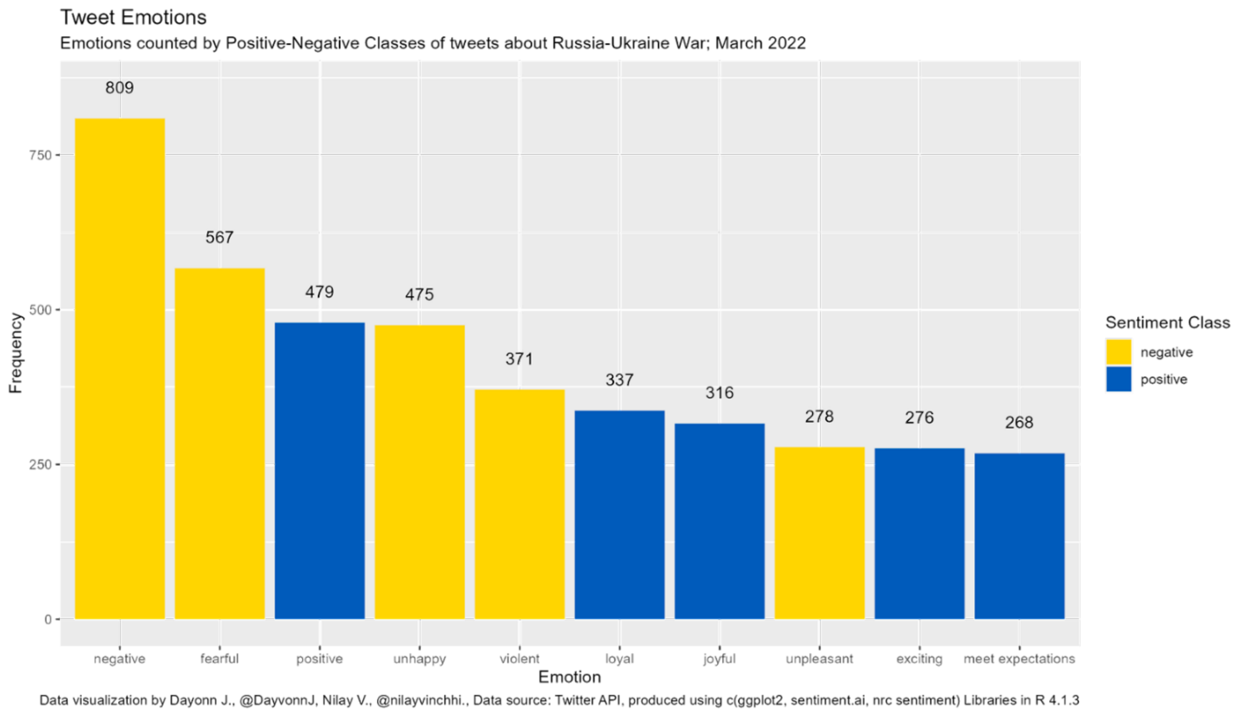

Figure 9: Sentiment Analysis using bing lexicon by Count for tweets in Ukraine about the Ukraine-Russia war 2022