



# RUSSIAN INVASION OF UKRAINE

Sentiment Analysis

## ABSTRACT

Summary of suggestions to NGOs and the government of Ukraine on topics that would help improve the image of Ukraine on an international level. This analysis was conducted on social media and suggestions derived using various analytical techniques.

Group 5

MIE1624

# Table of Contents

---

1. Sentiment Modelling .....	2
1.1. Data Collection.....	2
1.2. Data Pre-processing and Feature Extraction .....	2
1.3. Training models and hyperparameter .....	2
1.4 Sentiment Prediction .....	2
2. Results and Discussion .....	3
2.1. Distribution of tweets.....	3
2.2. Major stories.....	4
2.2.1. War crimes.....	4
2.2.2. Story of Crimea .....	5
2.2.3. Unpaid wages .....	5
2.2.4. NGOs leaving Ukraine .....	<b>Error! Bookmark not defined.</b>
2.2.5. India and China call for negotiations .....	6
2.2.6. NGOs trying to convince Russians to be democratic.....	<b>Error! Bookmark not defined.</b>
3. Conclusion.....	7
4. References .....	7

# 1. Sentiment Modelling

---

Sentiment analysis is a powerful technique that can be used to understand the general sentiment of large groups, much deeper than just hashtags trend, from the raw data that has been uploaded by users (Drus & Khalid, 2019). It has been applied here to gauge the overall sentiment of people regarding the war between Russia and Ukraine on various social media platforms because of its farther reach than conventional media forms (Xu, 2020). This analysis can help us make recommendations to NGOs working for Ukraine and the Ukrainian government. Sentiment in public is of significant importance in democracies and help shape the decisions taken by politicians regarding any foreign policies.

Using natural language processing (NLP) and Information retrieval (IR), we have derived some suggestions that would help in improving the image of Ukraine on international platforms and work on issues that most relevant and in trend. There are some topics that needs revision while some that needs to be addressed.

## 1.1. Data Collection

Various scrapers were used to scrape data from major social media platforms like Twitter, reddit and facebook to create a representative corpus. Tweets and its relevant metadata were scraped that were related to the topic in hand. Hashtags are one of the primary ways people socialize (Laucuka, 2018) and therefore make the perfect way to target this data collection.

## 1.2. Data Pre-processing and Feature Extraction

Data pre-processing is crucial for extracting interesting and non-trivial knowledge from unstructured text data. Retweets were removed to avoid duplicating analysis. Text documents are cleaned by first converting them into lower case, followed by removing URLs, custom designed stop words are removed from the text data, numeric characters and punctuations are also removed and the words were stemmed into its root form.

The cleaned text data is then transformed into numbers using TF-IDF vectorizer.

## 1.3. Training models and hyperparameter

Four different classification algorithm Logistic Regression, SVM, Random Forest and XGBoost are trained and tuned on the provided training data. The train data set is unbalanced, therefore F1 score which takes into account both precision and recall is used for tuning hyperparameters.

XGBoost performed best and is selected for predicting sentiments of the collected data set.

# 2. Sentiment Prediction

---

Sentiment prediction for the collected data is done by using 4 different models. They are AWS comprehend, Vader Sentiment Analyzer, NLTK Text Blob and XGBoost. XGBoost is trained and tuned on the provided training data and rest are pretrained model. Final sentiment value is assigned to each document by majority voting from the four predictions.

### 3. Results and Discussion

#### 3.1. Distribution of tweets

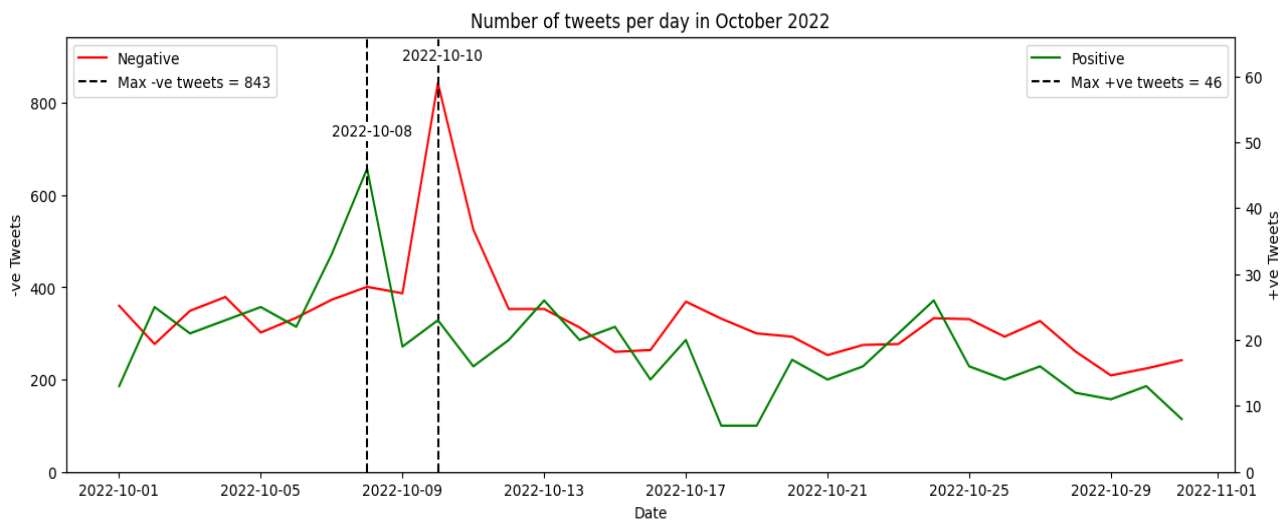


Figure 1: Number of Tweets Per Day In October

One of the techniques used in understanding the topics that drive sentiment towards the Russia-Ukraine conflict is by plotting the number tweets made during each day of the month in October. Figure 1 shows an example of such a plot. One peak can be seen for both tweets with positive sentiment and tweets with negative sentiment respectively. These peaks have been marked on the figure.

By first looking at the positive sentiment peak, news articles relating to the Russia-Ukraine conflict written on the day that the peak occurred were searched to see if any major events occurred. It was then found that on 8<sup>th</sup> of October 2022, there was an explosion on a bridge in Crimea. CNN reported that it was “the only direct

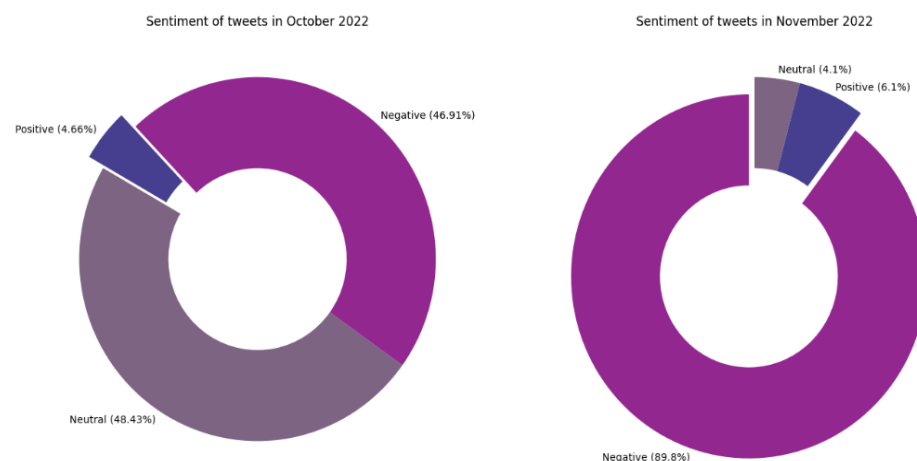


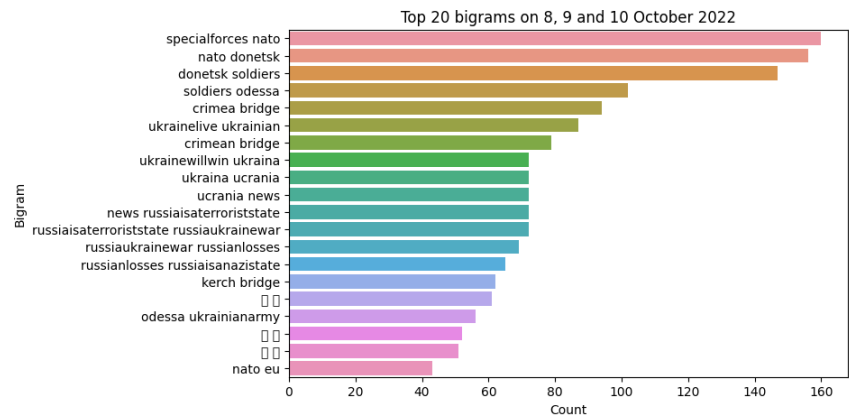
Figure 2: Distribution of sentiment for Oct and Nov

road and rail connection between annexed Crimea and mainland Russia” (Woodyatt, Vogt, & Meyer, 2022). Understanding that the explosion posed as a barrier for entry into Ukraine by Russian entities, it can be assumed that positive sentiment was shown by twitter users who support Ukraine.

A similar analysis can be done for the negative sentiment peak. The peak occurred on the 10<sup>th</sup> October 2022. CNN reported that on this day “at least 14 people were killed in Russian strikes across Ukraine” (Subramaniam, Haq, Upright, & Sangal, 2022). The news agency also stated that global leaders “condemned the Russian attacks and

vowed to continue to support Ukraine in its war efforts". The peak in negative sentiment suggests that a general dissent by Twitter users towards the events that occurred on the day.

Finally, by looking at the scales of the vertical axes in Figure 1, it can be seen that there was a much a larger proportion of positive tweets compared to negative ones, which is quite common for opinions expressed in real time (Garcia, 2020). This may also suggest a widespread feel of pessimism towards the Russia-Ukraine conflict.



## 3.2. Major stories

Using various analytics methods, we extracted the most relevant topics from the data. Following are the most relevant topics that we suggest need to be talked about to maintain a good rapport with international committees.

### 3.2.1. War crimes

The United Nations Human Rights Council created the Independent International Commission of Inquiry on Ukraine to investigate violations and abuses of human rights, violations of international humanitarian law, and related crimes that may have been committed due to the aggression by the Russian Federation against Ukraine. 18 October 2022, based on its investigations of the events in Kyiv, Chernihiv, Kharkiv, and Sumy regions in late February and March 2022, the Independent International Commission of Inquiry has found reasonable grounds to conclude that an array of war crimes, human rights violations, and international humanitarian law have been committed in Ukraine. The UN Commission documented patterns of summary executions, unlawful confinement, torture, ill-treatment, rape, and other sexual violence committed in areas occupied by Russian armed forces across the four regions (Kyiv, Chernihiv, Kharkiv, and Sumy) on which it focused. Sexual violence has affected victims of all ages. Family members, including children, were sometimes forced to witness the crimes. As of October, OHCHR had recorded 6,306 civilians killed and 9,602 wounded in all of Ukraine since 24 February 2022, where millions had lost homes and livelihoods and were forced to flee (Independent International Commission of Inquiry on Ukraine, 2022).

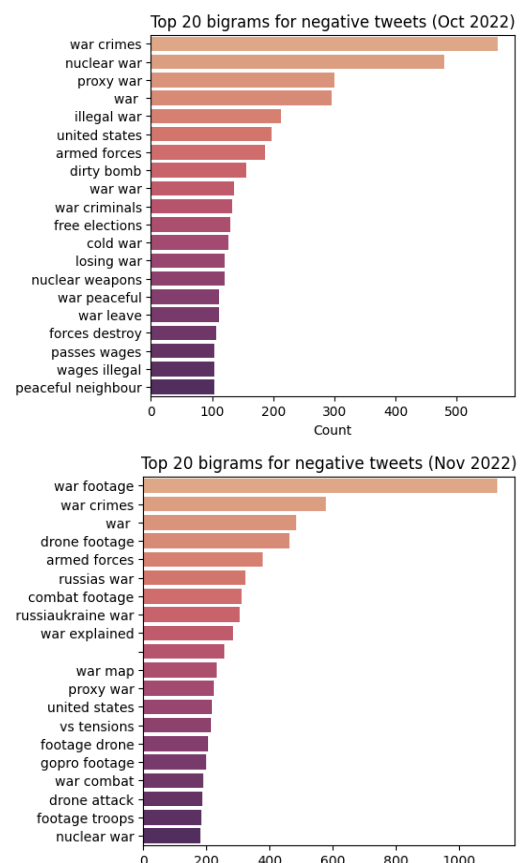


Figure 3: Top 20 bigrams for negative tweets with war crimes amongst the top

### 3.2.2. Story of Crimea

Russia annexed Crimea in 2014, following the Maidan Revolution and faced heavy condemnation internationally. Crimea consisted of majority Russian ethnic group, (State Statistics Committee of Ukraine, 2001) and was very much under the control of Russia, with its clocks shifted to Moscow time and Russian Rubles in circulation. Ukraine had

in response terminated the water supply to Crimea from the Dnieper River (Garrood, 2020) which put a lot of strain on Russia to maintain livelihood for the population there. Capturing Kherson during its invasion allowed them to destroy the dam and return the water supply, but the current situation does not look good for them, as Ukraine managed to recapture Kherson (aljazeera, 2022). We can see in our analysis that the tweets mentioning Kherson during the time it was being recaptured by Ukraine had significantly less negativity than

our overall average, with only 33% negative compared to the overall 47%. Since the topic at hand is about war, we expect the general sentiment to be negative, but this could mean that people are in favour of Kherson being in hands of Ukraine. We would suggest NGOs and the government of Ukraine to work on ground with people there and report more stories about them to keep the general sentiment of people in its favour.

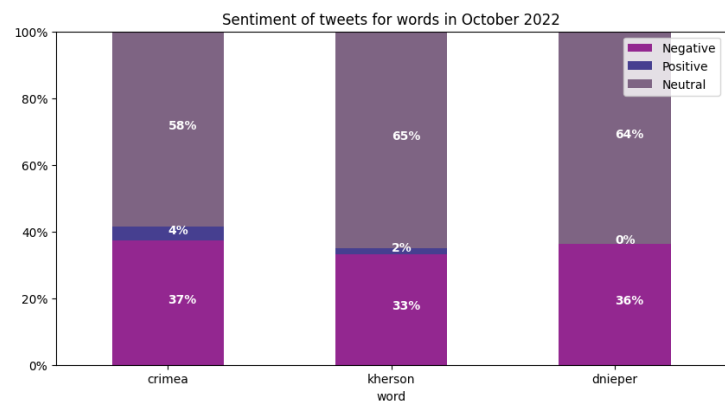


Figure 4: Sentiment Breakdown of Tweets that include words related to Crimea

### 3.2.3. Unpaid wages

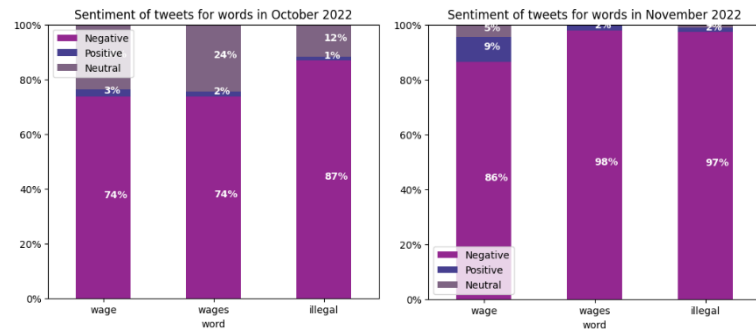


Figure 5: Wordcloud from one of the topics derived from LDA analysis of negative tweets

Figure 4 shows a wordcloud visualization created using the most common tri-grams of negative sentiment from one of the topics generated using Latent Dirichlet Allocation (LDA). By taking a closer look at the visualization, it can be noted that the words, “wages” and “illegal” occur frequently. After conducting research pertaining to a negative sentiment towards illegal wages during the Russia-Ukraine conflict, news outlets such as the Hindustan Times reported of unpaid soldiers protesting against Vladimir Putin’s mobilization of the Russian army (Soni, 2022). The news outlet continued to say that over one hundred conscripted Russian soldiers were promised 195,000 Rubles



but never received their compensation. Not only were they left uncompensated, but some soldiers reported that they were not allowed to visit relatives and were also declined leave (Soni, 2022). Since the tweets used in the LDA analysis were all deemed to be of negative sentiment by our models, it shows that there exists negative sentiment towards the Russia-Ukraine within Russia itself.

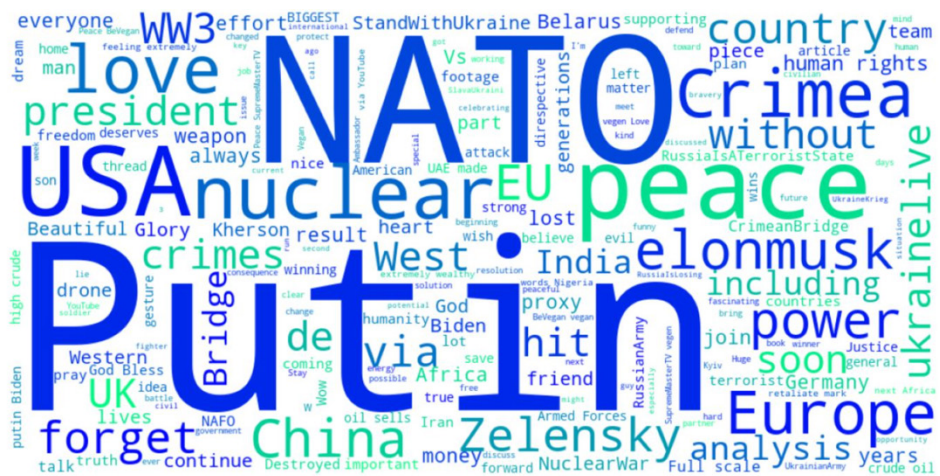


**Figure 6: Split of sentiments of tweets with wage related words**

### 3.2.4. India and China call for negotiations

Figure 4 presents a wordcloud visualization that was created using tweets with positive sentiment. During the early stages of the conflict countries such as India and China did not oppose Russia's decision to invade Ukraine. India was initially reluctant to condemn Russia (Miller & Jordan, 2022). Understanding that in general, there has been negative sentiment towards the Russia-Ukraine conflict, it initially seemed peculiar that India and China were both

largely present in the wordcloud visualization. Subsequently, it was found that India and China had recently changed their stance on the conflict. The Guardian reported that both countries started to “call for negotiations to end the war” (The Guardian, 2022). This may explain the positive sentiment towards the two countries with respect to their involvement in the Russia-Ukraine conflict.



**Figure 7: Wordcloud made from tweets with positive sentiment**

### 3.3 Recommendations to Ukraine Government and NGOs

Throughout the duration of the Russia-Ukraine conflict, there has been an abundance of fake news spreading misinformation about the war. This news has been spread by both the pro-Russia and Pro-Ukraine sides (Holroyd, 2022). Such fake news spreads rapidly across the population and can be detrimental to how Ukraine's image on the conflict is perceived. Our suggestion is to conduct sentiment analysis on popular websites where misinformation is spread such as twitter and Reddit. When examples of posts with large negative or positive sentiment are detected, a bot can be deployed to comment on the post and suggest reputable sources for users to read. An example of this can be seen in Figure 8. By educating the public about what is truly happening during the conflict, we believe that it will help in preserving Ukraine's presence and image. The bot's functionality can also be extended to link to NGO initiatives that help the people of Ukraine in their constant struggle facing the Russia-Ukraine conflict.

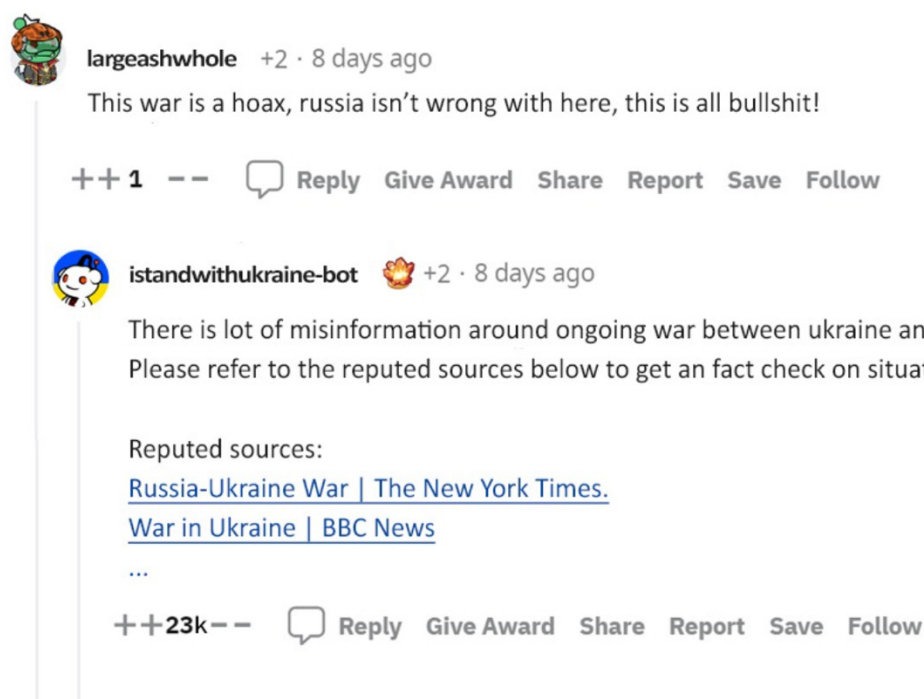


Figure 8: Example of employing bots

## 4. Conclusion

This report has outlined the use of sentiment analysis to understand the general public view of the Russia-Ukraine conflict. This was first done by training four sentiment analysing models with labelled data. Of all models, it was found that XGBoost performed the best. This, along with pretrained AWS, Text Blob and Vader sentiment analysing models were used to detect the sentiment of twitter posts written about the conflict between Russia and Ukraine using a majority voting scheme. Web scraping was done to gather all the twitter data. With the results of the sentiment analysis and the twitter data, tools such as LDA were used to understand the reasons behind positive and negative sentiment towards the Russia-Ukraine conflict. Some major stories that drove sentiment were about war crimes committed by Russia, the siege of Crimea, unpaid conscripted Russian soldiers, and India and China changing their stance on the conflict. Finally, a recommendation was made to employ bots on popular platforms where misinformation is spread. The bots would comment on posts and suggest reputable sources to safeguard against user's reacting to information that is not true.

## 5. References

- aljazeera. (2022, November 12). *'Kherson ours': Ukraine celebrates after Russian retreat*. Retrieved from aljazeera: <https://www.aljazeera.com/news/2022/11/12/ukrainians-celebrate-as-russia-withdraws-from-kherson>
- Drus, Z., & Khalid, H. (2019). Sentiment Analysis in Social Media and Its Application: Systematic Literature Review. *Procedia Computer Science*, 161, 707-714. doi:<https://doi.org/10.1016/j.procs.2019.11.174>
- Garcia, M. B. (2020). Sentiment Analysis of Tweets on Coronavirus Disease 2019 (COVID-19) Pandemic from Metro Manila, Philippines. *Cybernetics and Information Technologies*, 20(4), 141-155. doi:10.2478/cait-2020-0052



- Garrood, M. (2020, December 12). *Crimeans have tap water only six hours a day as all Russian attempts to hydrate occupied peninsula fail*. Retrieved 2022, from euromaidanpress: <https://euromaidanpress.com/2020/12/17/water-severely-rationed-in-crimean-cities-after-all-attempts-to-solve-crisis-fail/>
- Holroyd, M. (2022, 08 24). *Ukraine war: Five of the most viral misinformation posts and false claims since the conflict began*. Retrieved from euronews: <https://www.euronews.com/my-europe/2022/08/24/ukraine-war-five-of-the-most-viral-misinformation-posts-and-false-claims-since-the-conflict>
- Independent International Commission of Inquiry on Ukraine. (2022). *A/77/533: Independent International Commission of Inquiry on Ukraine - Note by the Secretary-General*. United Nations, General Assembly. Independent International Commission of Inquiry on Ukraine - Note by the Secretary-General. Retrieved December 2022, from <https://www.ohchr.org/en/documents/reports/a77533-independent-international-commission-inquiry-ukraine-note-secretary>
- Laucuka, A. (2018, July 19). Communicative Functions of Hashtags. *Economics and Culture*, 15(1), 56 - 62. doi:10.2478/jec-2018-0006
- Miller, M. C., & Jordan, Z. (2022, 05 10). *India Is Reluctant to Condemn Russia. Its History with China Looms Large*. Retrieved from Barron's: <https://www.barrons.com/articles/india-is-reluctant-to-condemn-russia-its-history-with-china-looms-large-51652132357>
- Noldofinve. (2014, December 9). *The Army*. Retrieved December 2, 2022, from deviantart: <https://www.deviantart.com/noldofinve/art/The-Army-499115251>
- Soni, M. (2022, 11 4). *Video: Unpaid soldiers protest against Vladimir Putin's mobilisation*. Retrieved from Hindustan Times: <https://www.hindustantimes.com/world-news/russia-ukraine-war-vladimir-putin-news-ukraine-war-video-unpaid-soldiers-protest-against-vladimir-putin-s-mobilisation-101667546766500.html>
- State Statistics Committee of Ukraine. (2001). *National composition of population*. Retrieved January 05, 2008, from ukrcensus.gov: <https://web.archive.org/web/20080105063616/http://www.ukrcensus.gov.ua/eng/results/general/nationality/>
- Subramaniam, T., Haq, S. N., Upright, E., & Sangal, A. (2022, October 10). *October 10, 2022 Russia-Ukraine news*. Retrieved from CNN: <https://www.cnn.com/europe/live-news/russia-ukraine-war-news-10-10-22/index.html>
- The Guardian. (2022, 09 24). *Russia's allies China and India call for negotiations to end Ukraine war*. Retrieved from The Guardian: <https://www.theguardian.com/world/2022/sep/24/russias-allies-china-and-india-call-for-negotiations-to-end-ukraine-war>
- Woodyatt, A., Vogt, A., & Meyer, M. (2022, October 8). *October 8, 2022 Russia-Ukraine news*. Retrieved from CNN: <https://www.cnn.com/europe/live-news/russia-ukraine-war-news-10-08-22/index.html>
- Xu, J. (2020). Does the medium matter? A meta-analysis on using social media vs. traditional media in crisis communication. *Public Relations Review*, 46(4), 101947. doi:<https://doi.org/10.1016/j.pubrev.2020.101947>