Table of Contents

1. Problem Description and Background

1.1. Description

In February 2022, a full-scale invasion of Ukraine was launched by Russian President Vladimir Putin. The rationale behind the Russian aggression was predicated on unfounded allegations that the Ukrainian government was under the control of neo-Nazi factions that were conducting ethnic cleansing against Russian minorities. This was accompanied by challenges to Ukraine's legitimacy as a sovereign state.

We are witnessing Russian war crimes unfolding in front of our eyes and Ukrainian women and children have been forced to leave their country. Europe is currently experiencing its biggest refugee crisis since World War II due to the ongoing war. The number of fatalities is rising rapidly, and there is a substantial risk of nuclear escalation from the Russian government as the conflict persists. To aid Ukraine in defending itself against Russian invasion, numerous countries around the globe have provided both humanitarian and military assistance to the country.

The Ukrainian government has hired Group 5 from the MIE1624 “Introduction to Data Science and Analytics” course to conduct a comprehensive evaluation of international perception in the context of the ongoing war. The group's objective is to recommend actions that the Ukrainian government and its media outlets can take to improve Ukraine's image and increase its prominence on the global stage. This effort seeks to enable the Ukrainian government to effectively communicate their genuine message to the world's populace.

1.2. Solution Strategy

The team has developed a sophisticated model capable of comprehending and analyzing the general public's sentiment from Twitter concerning the Russian invasion of Ukraine. By meticulously scrutinizing sentiment trends in conjunction with significant wartime events, we suggest strategies that could be strategically leveraged to maintain Ukraine's honest reputation on the global stage and increase the reach of the government’s message. The resulting insights will inform the development of proactive communication tactics that will enable the Ukrainian government and its media outlets to effectively and accurately shape international opinion, positively impact how Ukraine is being perceived, and demonstrate its unwavering commitment to sovereignty, stability, and democracy.

2. Methodology and Analysis

2.1. Sentiment Model

Machine learning (ML) models are frequently used for sentiment analysis. To conduct sentiment analysis, the first step is training a machine learning model on a large dataset of labeled text examples, where the labels are the associated sentiment of texts. (e.g., positive, negative, neutral). The trained model is then used to predict the sentiment of new, unseen text data.

To carry out sentiment analysis, we utilize a comprehensive dataset of tweets labeled with either a Positive or Negative sentiment as the ground truth. For our analysis, we focus on the binary classification of the sentiment as it enables us to more effectively analyze the events and strategies employed by different stakeholders of the Ukraine-Russia war. This dataset is used to train a variety of classification models, including Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, and Decision tree algorithms. To transform the text into numerical features, we use the Term Frequency Inverse Document Frequency (TF-IDF) technique. A fivefold cross-validation is employed to identify the best set of hyperparameters for the models. Finally, we implement an ensemble model that combines the predictive strengths of each model to be able to classify the sentiment of a given tweet.

Table 1: An overview of the achieved test accuracy by our Machine Learning algorithms

Model	**Multinomial NAIVE BAYES**	**Random Forest**	**Linear Support Vector Machine**	**Logistic Regression**
Training Accuracy	93.5%	95.7%	96.4%	96.6%
Test Accuracy	92.5%	92.9%	95.7%	95.8%

2.2. Adding Context to Model

In the preceding section, our team successfully trained an ensemble model on a corpus of tweets devoid of any references to the ongoing conflict between Russia and Ukraine. Nonetheless, our ultimate objective is to perform sentiment analysis on tweets specifically concerning the war, with a view to comprehending the global sentiment surrounding Russia's aggression and invasion of Ukraine. Hence, to achieve this goal, we must train our model on a contextualized dataset, one that incorporates the necessary nuances and intricacies of the conflict.

In order to create a dataset that includes both pro and anti-Russian aggression tweets, we sourced two Kaggle datasets, containing tweets about the Ukraine-Russia war. To identify tweets that are anti-Russian aggression, we utilized a dataset of tweets with the hashtag "StandwithUkraine," which is indicative of anti-Russian sentiment. We filtered negative tweets from this dataset by examining tweets containing words such as "crime," "violence," "invasion," "evil," "genocide," and "death," and labeled them as negative sentiment tweets.

To identify tweets in support of Russian aggression towards Ukraine, we used a Kaggle dataset of tweets about the Russia-Ukraine war, and to find pro-Russian war tweets, we searched for tweets containing hashtags such as "standwithrussia," "istandwithrussia," "standwithputin," "istandwithputin," and labeled them as positive sentiment tweets towards the war if they included some keywords with high positive sentiment such as "Success", Victory", "Happy", "Great","bravery ". We then combined this dataset, which includes both positive and negative tweets about the Russia-Ukraine war, with a subset of the original sentiment dataset to train the final version of our ensemble model.

2.3. Best Model Unbiased Performance

After training our ensemble model on the training data which contains the original sentiment dataset plus the developed sentiment data about the Russian-Ukraine war described previously, we tested the ensemble model performance on the Kaggle dataset that contains the "russiaukrainewar" hashtag. Since this dataset lacks ground truth sentiment labels, we used the Google Natural Language API to determine sentiment scores for each tweet and subsequently classified them as either having positive or negative sentiment. Our ensemble model's performance on this test data, which includes augmented labels from the Google NLP API, was approximately 85%. This suggests that our model's performance is at least on par with that of the Google NLP API.

3. Assessing International Public Sentiment

To gauge international sentiment regarding the Russian invasion of Ukraine, we analyzed tweets containing hashtags such as "Ukrainewar," "Russianukrainewar," "StandWithRussia," and "StandWithUkraine," among others, from Feb 2022 to March 2023. We ensured unbiasedness in the data extraction step by including hashtags supportive of each country to extract relevant tweets. For instance, when extracting tweets with "standwithzelensky," we also collected tweets with hashtag "standwithputin."

Figure 1 depicts the evolution of public sentiment towards the conflict, demonstrating that the majority (78%) of tweets have negative sentiment over the one-year period analyzed.

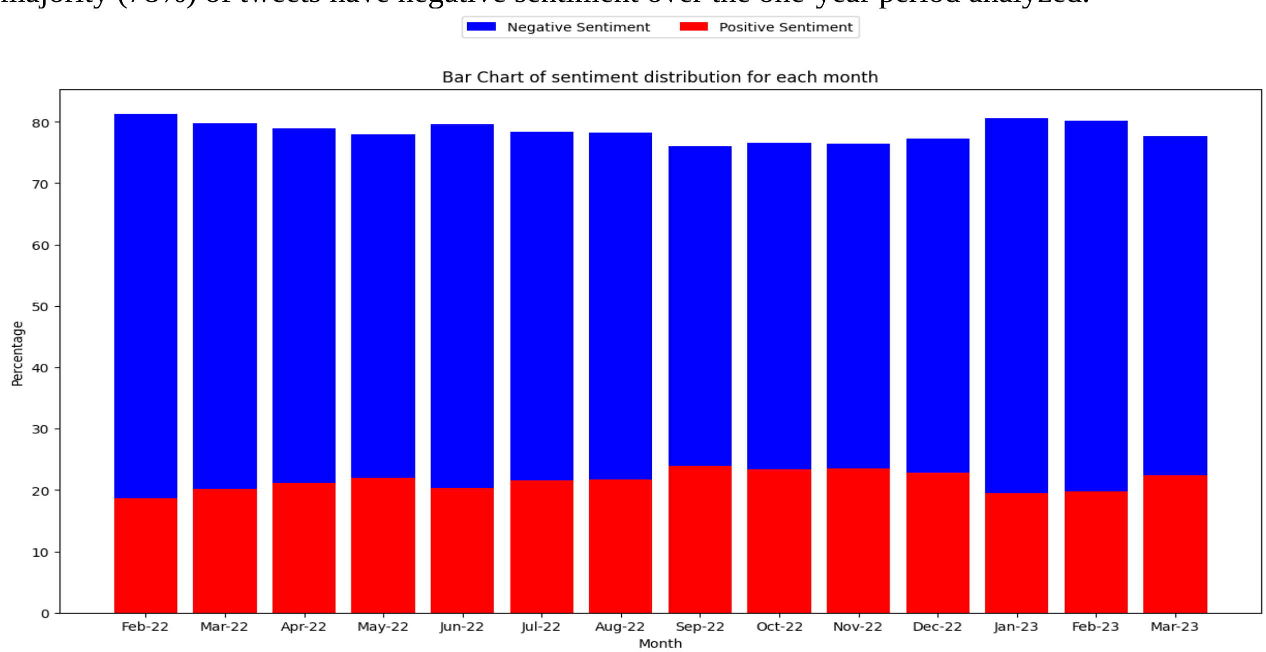

Figure 1: Monthly progression of sentiment since the start of war

Figures 2 and 3 display word clouds of commonly used terms in negative and positive sentiment tweets, respectively. As expected, these terms are highly related with the Russia-Ukraine war('Ukraine', 'Russia', 'Kiev', etc.), its key stakeholders ('Ukrainian', 'Zelensky', 'NATO', 'Putin', etc.), and the significant events('Keiv', 'Bakhmut').

Figure 2: Word Cloud from Pro-Ukrainian Tweets (or Tweets with Negative Sentiment)

Figure 3: Word Cloud from Pro-Russian Tweets (or Tweets with Positive Sentiment)

When developing our sentiment classification model, we hypothesized that negative sentiment tweets would condemn the Russian invasion of Ukraine, and positive sentiment tweets would support Russia. However, we found out that positive sentiment tweets may also support Ukraine and its people, while negative sentiment tweets could be negative propaganda produced by the Russian government against Ukraine.

Therefore, we manually analyzed 50 tweets from each positive and negative sentiment category to determine the percentage of tweets in support of Ukraine and critical of Russia versus the percentage of tweets being propaganda in support of Russia and Critical of Ukraine. Since the outbreak of the conflict, Russian propaganda has resorted to unprecedented measures to consistently mislead the international community regarding the events unfolding during the invasion of Ukraine. [2]

According to our analysis from the sampled tweets, 50% of positive sentiment tweets expressed support for Ukraine, while the other 50% were propaganda expressing support for Russia. On the other hand, 80% of negative sentiment tweets were critical of Russia and expressed support for Ukraine, while the other 20% could be classified as propaganda supporting Russia and criticizing Ukraine. Based on the proportion of tweets supporting Ukraine in each sentiment category, even after all of the Russian fake propaganda, we can see that Ukraine has the majority of support on the international stage.

4. Country Level Activity and Sentiment

Figure 4 depicts the volume of tweets and corresponding sentiment from various countries. The United States, the United Kingdom, Australia, and most European countries show the highest level of activity, with a predominance of negative sentiment in the majority of their tweets. Conversely, many countries such as China, South American countries, African countries, and the Middle Eastern countries demonstrate low levels of activity with regards to tweeting about the conflict.

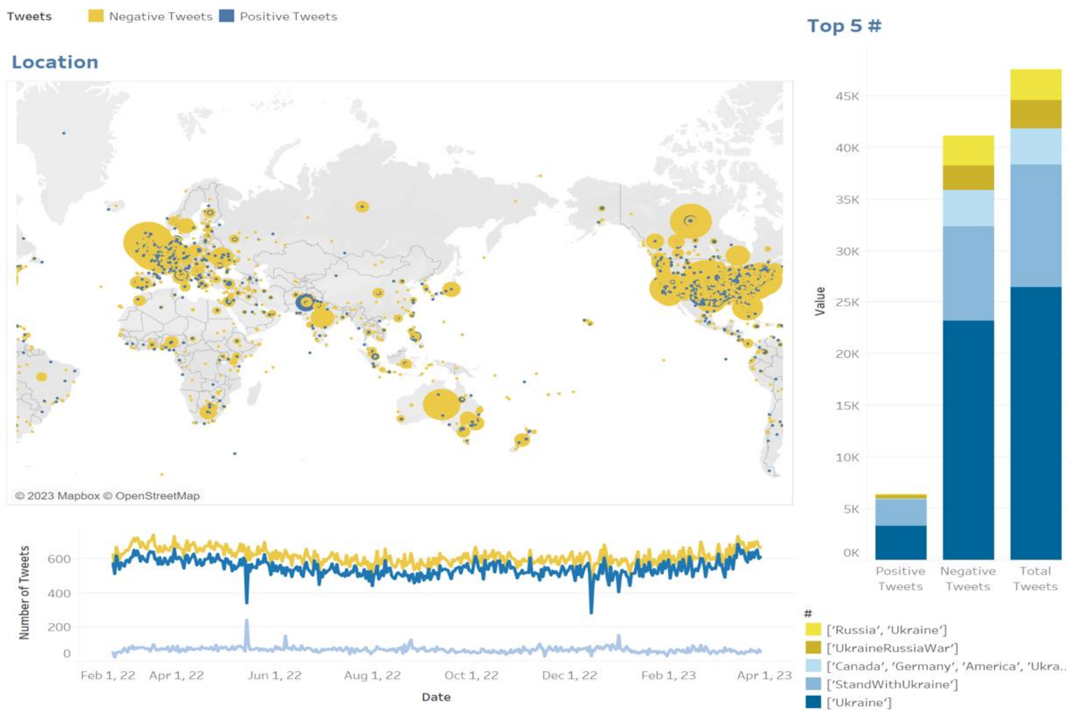

Figure 4: Sentiment and volume of tweets from each country

While language barriers could contribute to the lower levels of activity in these countries, it is crucial to acknowledge that many of the mentioned countries are dictatorships with close alliances to the Russian government [3]. However, this should not be taken as an indication that people in these dictatorships are in favor of the war in Ukraine, as limited access to the internet and social media are prevalent in these countries.

Overall, this visualization remains a critical representation of Ukraine's allies compared to countries led by dictatorships that support Russia. It can be used to develop strategies aimed at neutralizing other dictatorships' support for Russia while also focusing on delivering the accurate message about the situation to people living in these dictatorships.

5. International Public Sentiment & Europe Energy Index

In this section, we examine the relationship between the sentiment of tweets and a macroeconomic variable, “the harmonized index of consumer prices for energy in the Euro area” [4], which reflects the energy consumption of 19 European Union countries. We chose to investigate the consumer price index for the EU because of Europe's heavy dependence on energy imports from Russia, despite global warnings. Our analysis, as shown in Figure 5, reveals a negative correlation (-0.7) between the negative sentiment expressed in tweets and the rise in energy prices. The percentage change of the consumer energy index and negative and positive sentiment tweets are depicted in figure below. The correlation matrix between the percentage change of the consumer energy index and tweets sentiments can be found in Appendix Figure 1.

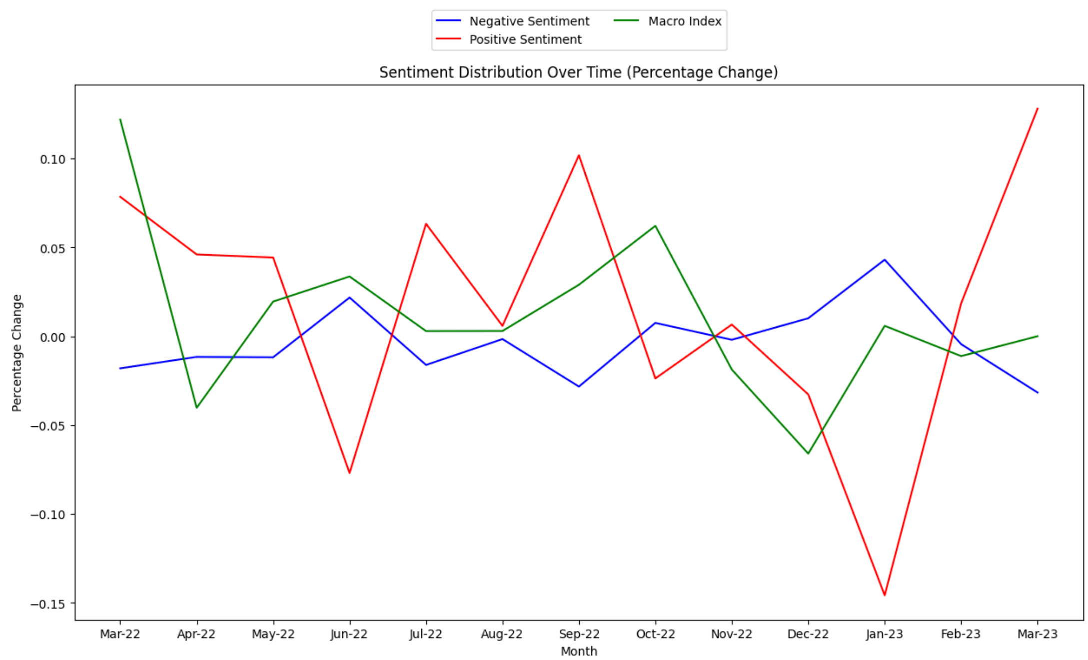

Figure 5: EU Energy Consumer Price Index and tweet sentiments monthly percentage change since the start of war

An intriguing observation is the negative correlation between the increase in energy prices and negative sentiment expressed in tweets. This suggests that public sentiment on the war is heavily influenced by events on the ground. It is imperative that European decision-makers recognize that the rise in fuel prices cannot be used as an excuse to slow the imposition of sanctions on Russian energy companies and entities [5]. The EU's continued dependence on Russian energy imports, instead of diversifying energy sources, must be immediately halted as divestment from Russian energy is crucial in cutting off funding for Putin's war crimes

6. Sentiment Vs. Traction

To scrutinize the potential influence of sentiment on a tweet's popularity, our team conducted an analysis of the correlation between tweet sentiment and traction. Traction is a metric that assesses a tweet's resonance by considering its total number of likes and retweets. By virtue of this measure, the more likes and retweets a tweet garners, the greater its traction, and the more probable it becomes to achieve viral status. Our investigation into this correlation can furnish the Ukrainian government and news outlets with critical insights that will empower them to craft messages strategically, optimizing their virality and amplifying their reach on the global stage.

Based on Figure 8, it appears that there is no correlation between tweet sentiment and traction. This shows that the international community are assessing the tweets message and information conveyed within them objectively and not engaging with or amplifying content based on personal biases. Hence, the Ukrainian outlet and media can have more confidence that people outside Ukraine are objectively analysing these tweets. This is crucial in the context of Russian propaganda, which uses emotional appeals and sensationalized content to spread false or misleading information. Hence by objective analysis the international community has shown that they resist these manipulative tactics from Russia and make informed decisions based on the facts presented. This fact also highlights the importance of responsible journalism and accurate reporting.

7. Twitter Sentiment During Major War Events

The table 1 in the appendix depicts chronicles major events that have occurred since the Russian invasion of Ukraine up until March 2023. We have categorized these events based on their alignment with or opposition to Ukraine's interests. By analyzing the sentiment of tweets surrounding these events (the day before, the day of, and the day after), we aim to assist the Ukrainian government in adjusting their messaging accordingly. Our research indicates that the change in sentiment is consistent with our previous findings, demonstrating the international community's support for Ukraine during important events in the war. For example, whenever a tragic event occurs within Ukraine, such as the surfacing of videos depicting Russian killings and brutal aggression, or on the other hand, a victory for Ukrainian people by reclaiming back their stolen territory, a noticeable change in negative sentiment is observed. Figure 7 depicts an example of such changes in sentiment in relation to some of the events in the Ukraine-Russia war. More examples of such events are provided in the appendix in figures 10 to16.

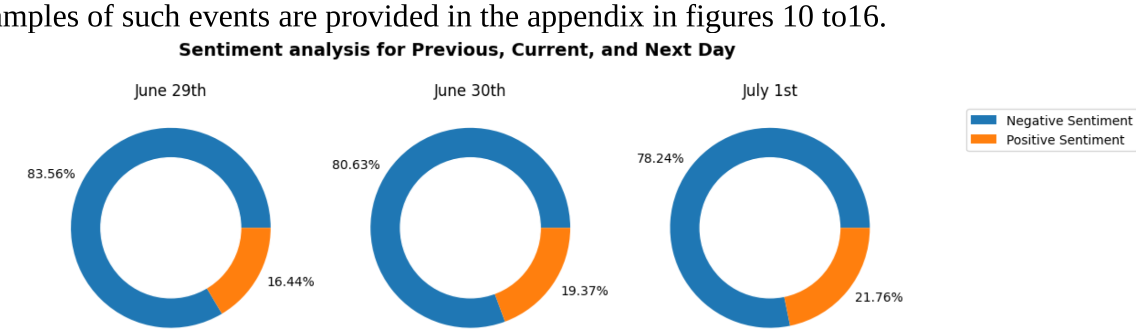

Figure 6: Ukraine fetes Russian pullback from strategic Snake Island outpost. [6]

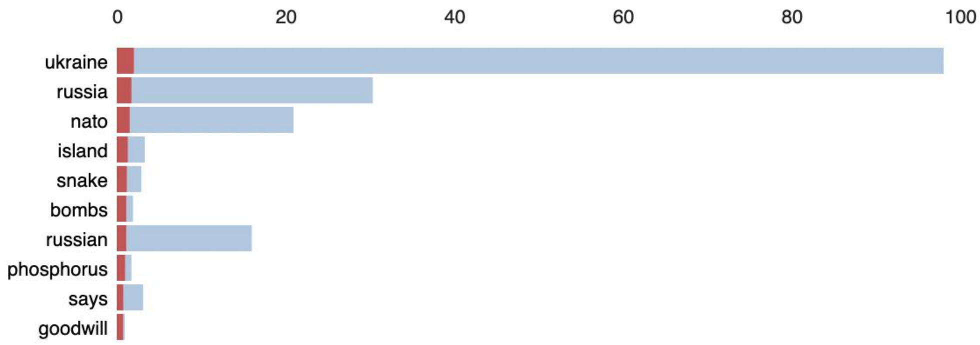

Figure 7: Most relevant terms for topic (red) total frequency (Blue)

8. Strategy Proposals (Recommendation)

Continuous assessment of sentiment in tweets related to the war, as well as their connection to major developments, is crucial for the Ukrainian government to adjust its messaging effectively. It is essential to identify key points that can frame a message for wider reach and virality.

Furthermore, to understand the level of support for Ukraine and the opposition to it, random sampling of tweets from each sentiment group should be conducted regularly. Establishing constant monitoring of tweet sentiments for any sudden shift in one sentiment direction, enables the detection of the event causing this shift. In case of a shift in balance in favor of Ukraine, the same identified reason and narrative should be adopted. On the other hand, if it is against Ukraine, the authorities should immediately take action to debunk and get ahead of the spread of possible tweets that are without any factual basis. This is crucial, especially if the shift in sentiment is due to new fabrication or propaganda from Russian outlets.

Previous analysis in this report has shown that during the energy price surge, most tweets were in support of Ukraine. Moreover, any criticism that was not directed toward Russia was aimed at European leaders who did not divest from Russian energy imports. Therefore, Ukraine should spread the message of boycotting Russian energy imports and imposing sanctions, without any concern about the potential loss of support from the international community due to the increase in energy prices.

9. Analysis Limitation

This section summarizes the limitations of our analysis that need to be considered when interpreting the results. It is crucial to address these limitations in future studies to obtain a more comprehensive understanding of the topic.

One limitation is that the analysis solely relies on Twitter, which may not accurately represent the global population's opinions. The sentiment model might also struggle to classify sentiments from other platforms, such as Facebook, Instagram, and LinkedIn, potentially leading to biased conclusions. Additionally, the study only examines English-language tweets, limiting the representation of non-English speaking countries' opinions and potentially hindering the accurate capture of international sentiment. Moreover, a more nuanced understanding of public opinion on the Russia-Ukraine conflict could be achieved by expanding the model to classify tweets into three categories, including neutral sentiment. Sentiment trend analysis should be approached with caution, taking into account the influence of propaganda and misinformation, particularly from state-sponsored sources like Russia, which could distort the results.

Increasing the computational capacity to train our model on larger datasets may help capture a broader and deeper understanding of sentiment patterns related to the topic. A more extensive sample size and comprehensive tweet mining methods, rather than relying on manually identified hashtags, could provide a more accurate and representative analysis. Furthermore, basic machine learning algorithms for sentiment analysis might not fully grasp the nuances and complexities of human language, leading to inaccuracies. Advanced algorithms, such as transformers, that account for more complex text structures could result in a more accurate and nuanced understanding of sentiment patterns.

In conclusion, these limitations should be acknowledged when interpreting the study's findings. While the results may offer valuable insights into sentiment patterns related to the topic, caution is advised when generalizing the findings beyond the study's limitations.

10. Conclusion

In conclusion, our sentiment analysis project has successfully evaluated the international perception of Ukraine during the ongoing Russian invasion. Through the use of machine learning algorithms and contextualized datasets, we have developed a sophisticated ensemble model capable of accurately predicting sentiment in tweets related to the conflict.

Our findings indicate that there is a substantial amount of negative sentiment towards Russia's aggression towards Ukraine, as evidenced by the abundance of tweets with negative sentiment towards the war. Additionally, our model suggests that strategic messaging surrounding topics such as Ukraine's commitment to sovereignty, stability, and democracy could help improve Ukraine's international reputation and increase its prominence on the global stage.

All in all, this project highlights the potential of sentiment analysis as a valuable tool for governments and media outlets in shaping international opinion during times of conflict. By leveraging machine learning algorithms and contextualized datasets, we can gain valuable insights into public sentiment and develop effective communication strategies that promote a more accurate understanding of complex geopolitical issues.

11. Appendix

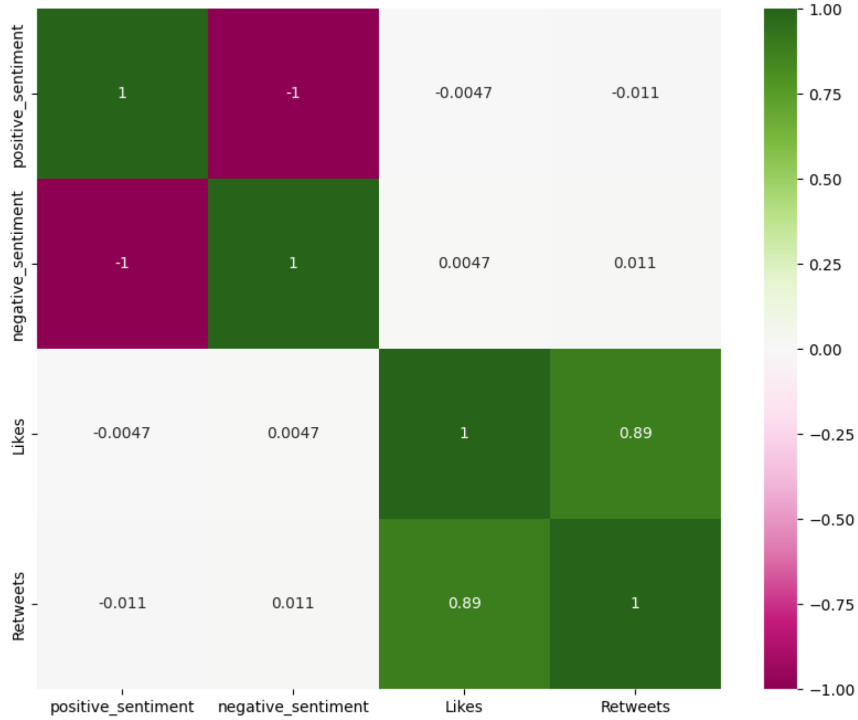

Figure 8: Correlation of sentiment and likes and retweets.

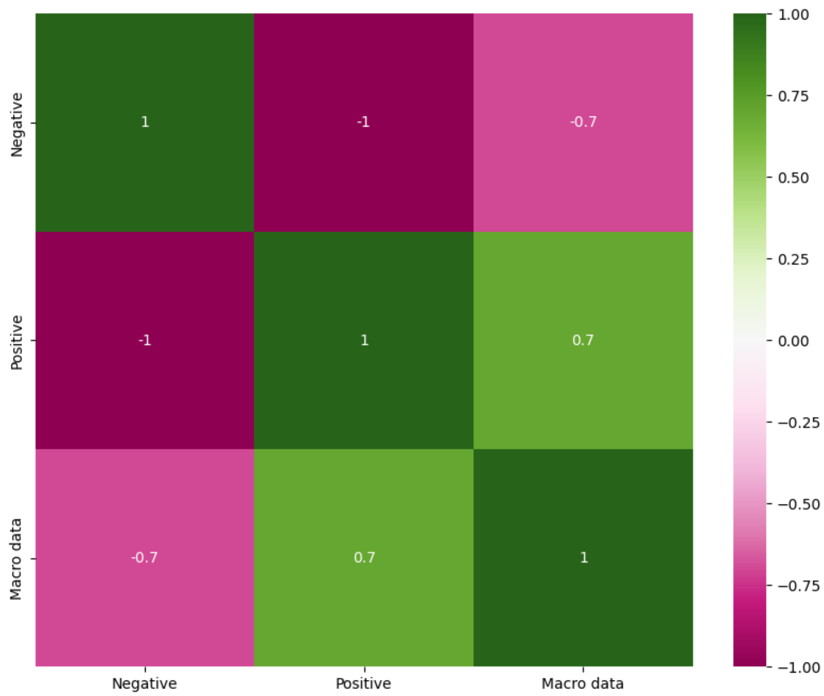

Figure 9: Correlation between tweet sentiment and The Harmonized Index of Consumer Prices: Energy for Euro

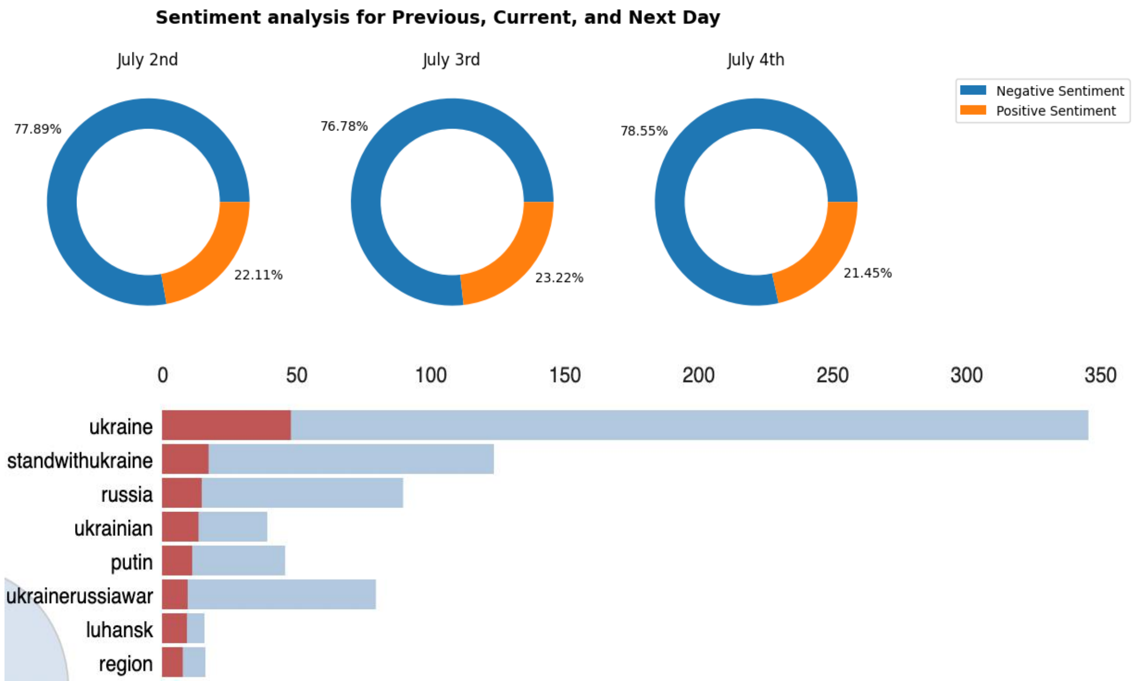

Figure 10: Last major Ukraine-controlled city in Luhansk region falls to Russia. [7] Sentiment change pie-plot and most relevant words with respect to the event.

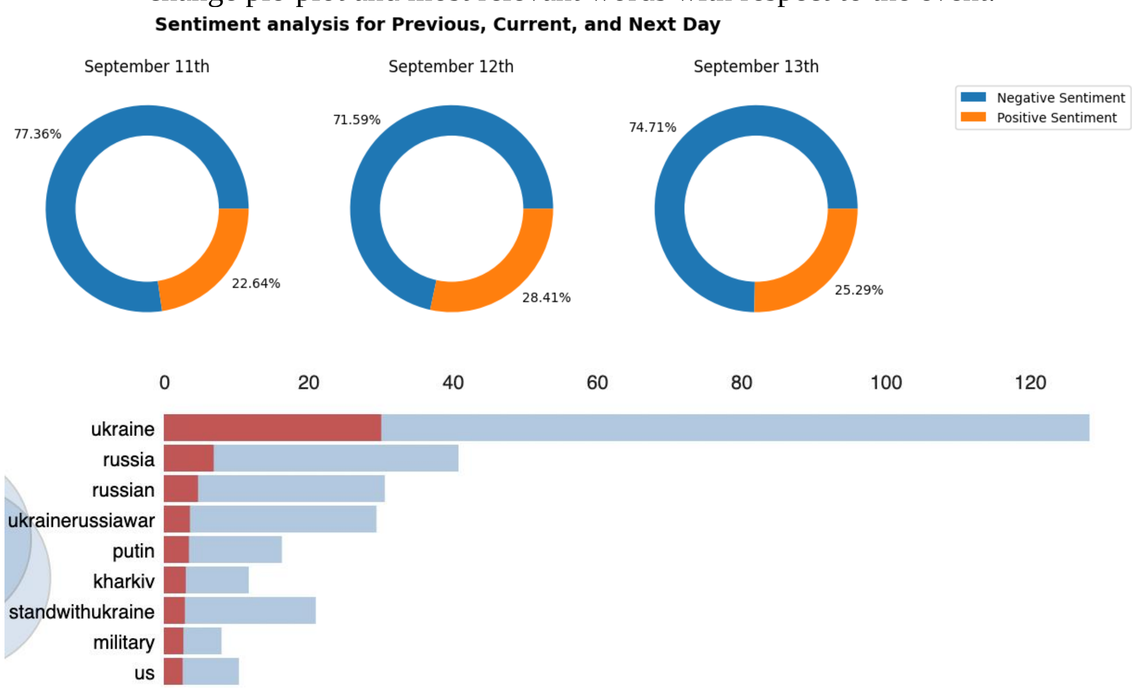

Figure 11: Ukraine reclaims more territory, reports capturing many POWs.[8] Sentiment change pie-plot and most relevant words with respect to the event.

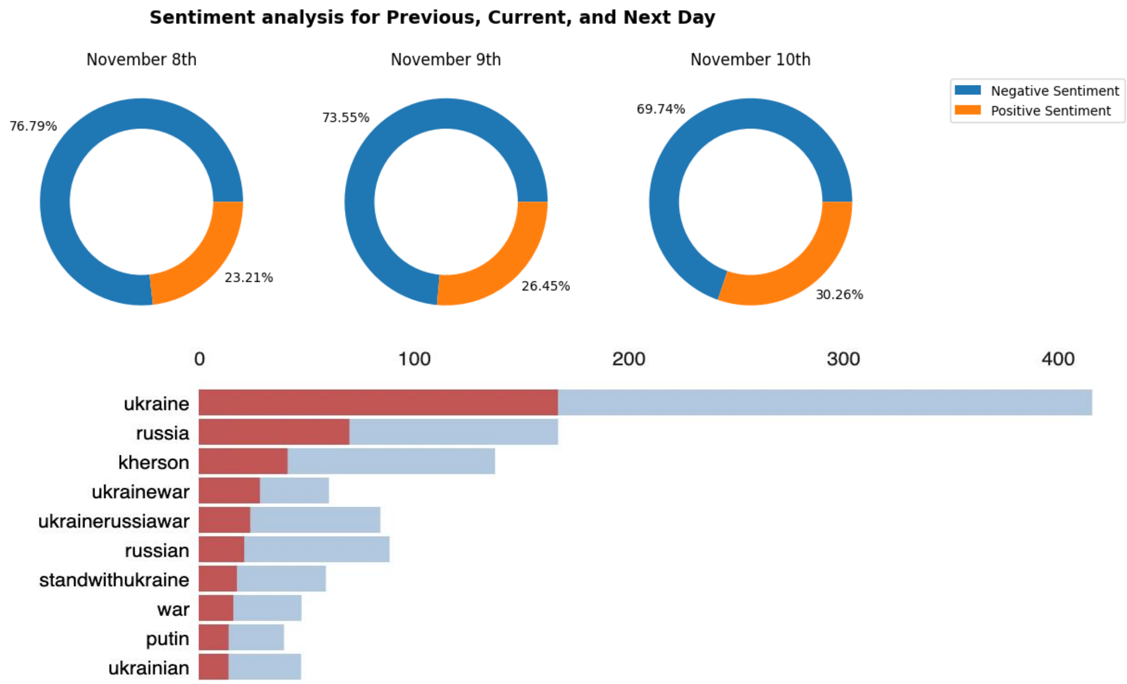

Figure 12: Russia says it's withdrawing from the key city of Kherson. [9] Sentiment change pie-plot and most relevant words with respect to the event.

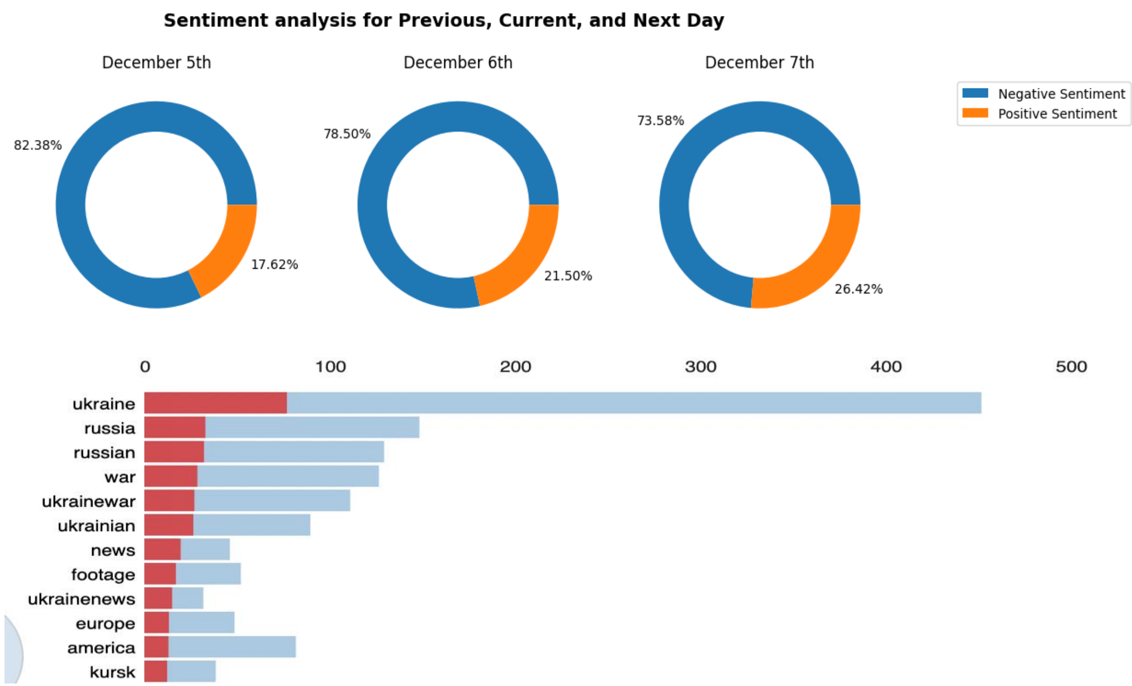

Figure 13: Ukraine strikes another Russian air base.[10] Sentiment change pie-plot and most relevant words with respect to the event.

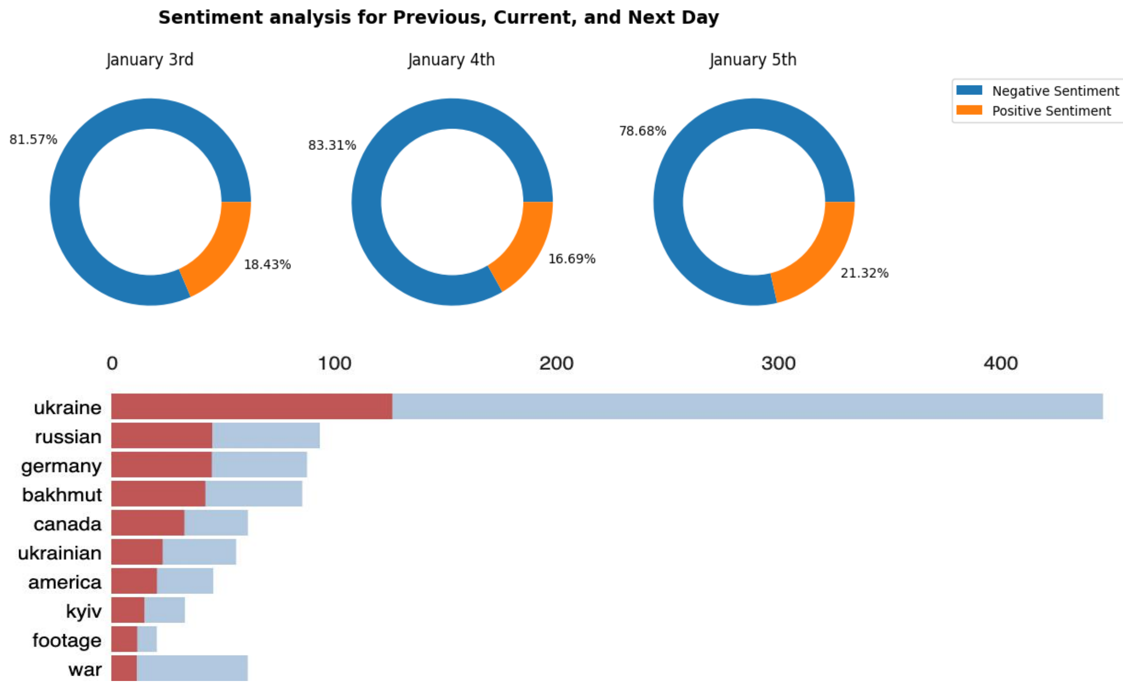

Figure 14: Russian military under scrutiny after 89 soldiers were killed in single Ukraine artillery attack. [11] Sentiment change pie-plot and most relevant words with respect to the event.

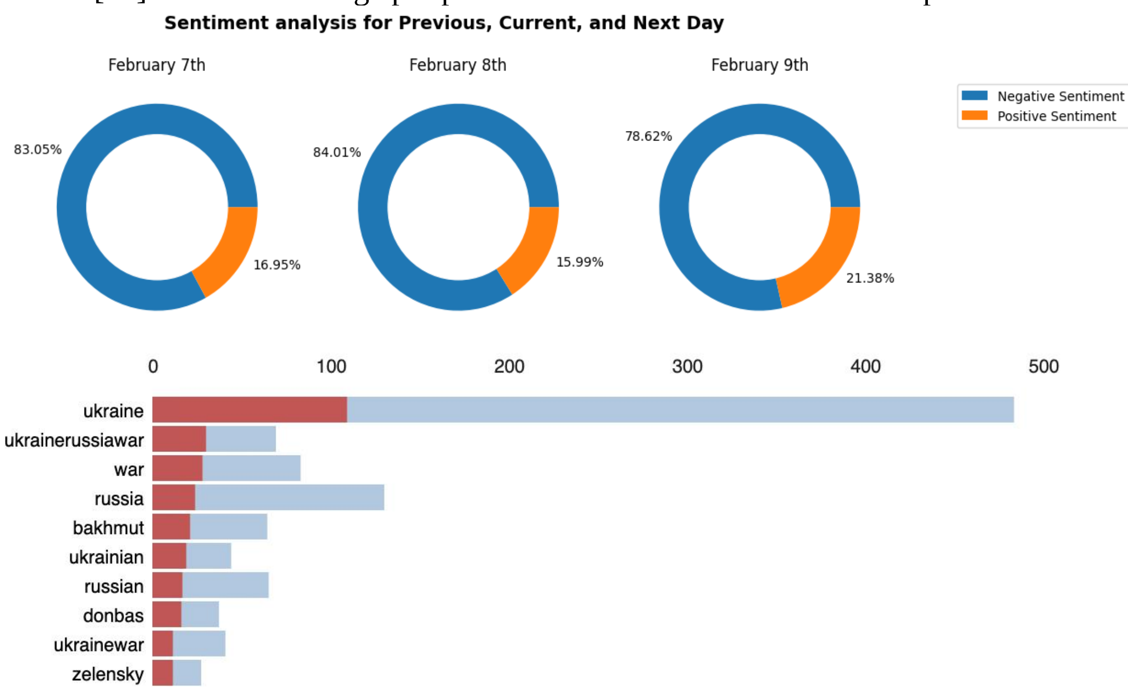

Figure 15: Zelensky Appeals for Fighter Jets. [12] Sentiment change pie-plot and most relevant words with respect to the event.

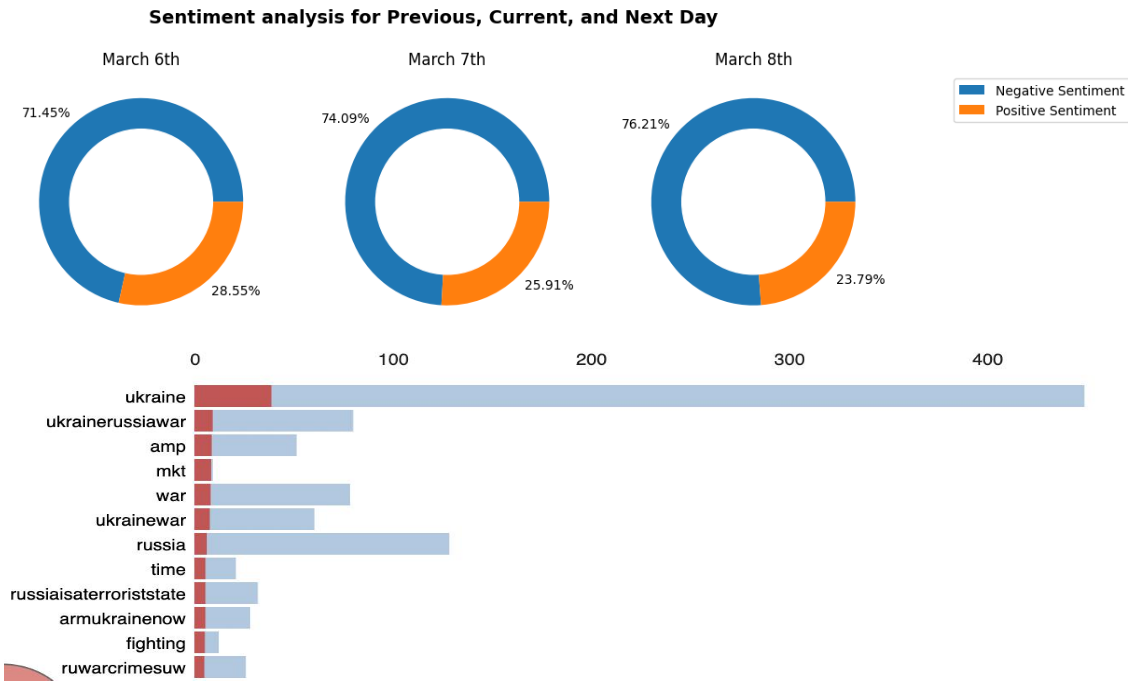

Figure 16: The video of a Ukrainian Soldier getting executed.[13] Sentiment changes pie-plot and most relevant words with respect to the event.

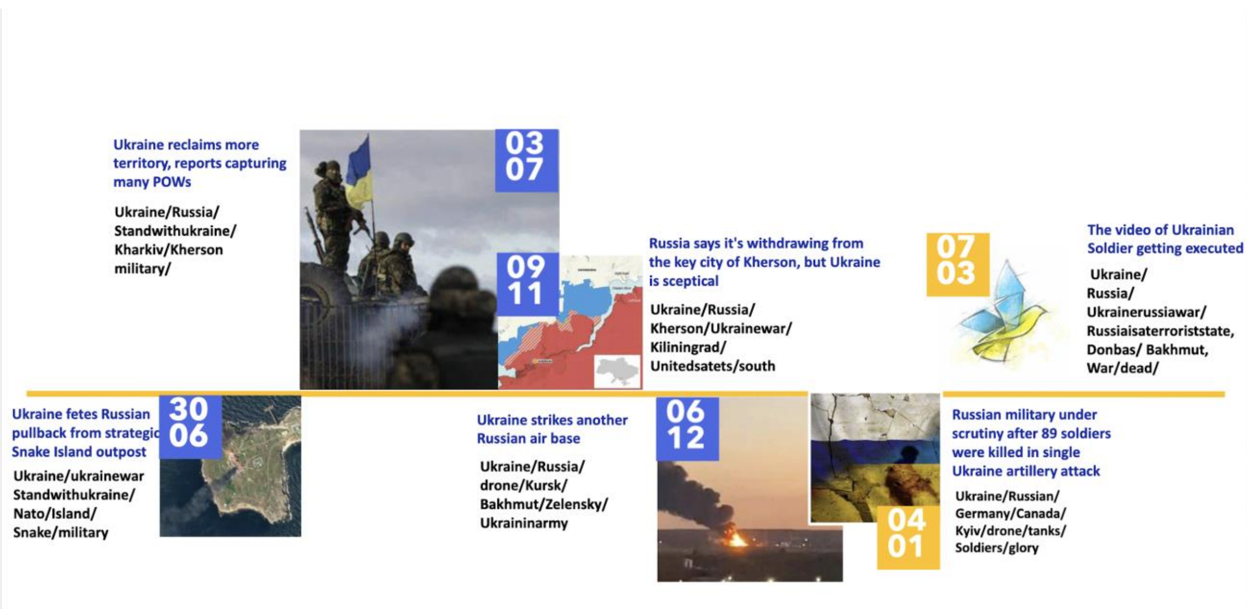

Figure 17: Chronological order of some of important events during the war and the associated key words

12. References

[1] https://time.com/6143645/how-to-stop-putin-invastion-ukraine/
[2] https://time.com/6257372/russia-ukraine-war-disinformation/
[3] https://www.aljazeera.com/news/2022/2/25/where-do-russias-allies-stand-as-western-powers-slam-moscow
[4] https://fred.stlouisfed.org/series/ENRGY0EZ19M086NEST#0
[5] https://www.politico.eu/article/ukraine-russia-business-sanctions-risk-oil-trading-ethical/
[6] https://www.reuters.com/world/europe/ukraine-says-russian-forces-have-withdrawn-snake-island-2022-06-30/
[7] https://www.politico.eu/article/ukraine-lysychansk-luhansk-region-russia-defeat/
[8] https://www.usnews.com/news/world/articles/2022-09-12/ukraine-keeps-initiative-claims-it-reached-russian-border
[9] https://www.npr.org/2022/11/09/1135513599/russia-withdraws-kherson-ukraine
[10] https://www.washingtonpost.com/world/2022/12/06/ukraine-drones-russian-airfield-attacks/
[11] https://www.pbs.org/newshour/world/cellphone-use-by-russian-soldiers-allowed-ukraine-to-target-troops-russian-military-says
[12] https://www.france24.com/en/europe/20230209-zelensky-repeats-demand-for-fighter-jets-on-france-uk-visits
[13] https://www.dailymail.co.uk/news/article-11827481/Horrific-video-shows-Ukrainian-POW-taking-one-drag-cigarette-executed.html