What Can Public Opinion Tell Us About the Russian War Against Ukraine in 2022? - Insights from Sentiment Analysis of Social Media Data

Table of Contents

1 Executive Summary

Russia's invasion on Ukraine started in February 2022 and has been the biggest threat to international peace and security in the 21st century. This consultant report uses public opinion expressed via social media to gain insights into the war to better assist the Ukraine government and NGOs in modifying its campaign strategy and coverage to receive support from a wider international audience. I have attempted to do so by applying sentiment analysis on a few biased and unbiased social media datasets spanning the period of the conflict. A model for Natural Language Processing Sentiment Analysis created by our group for this purpose is also compared against pre-existing Python libraries to test its performance. Several key sentiment drivers were identified based on the analysis of extreme sentiments, both negative and positive, to get a clear idea of how Ukraine is viewed in light of this invasion and recommendations are made to the government and international NGOs about steps that I believe will positively impact Ukraine's international presence and image.

2 Introduction

2.1 Background on Russia's War on Ukraine

Russian president Vladimir Putin made a shocking speech on February 21, 2022, to justify the "special military operations" that followed the next day. This speech included grievances like the long-simmering dispute over the expansion of the North Atlantic Treaty Organization (NATO) and post–Cold War security architecture in Europe while mostly focusing on questioning the legitimacy of Ukrainian identity and statehood themselves. Putin has long expressed his claim that distinct Ukrainian and Belarusian identities are the product of foreign manipulation and that, today, the West is following in the footsteps of Russia's imperial rivals in using Ukraine (and Belarus) as part of an "anti-Russia project". This claim suggests that Russia's ambitions extend beyond preventing Ukrainian NATO membership and encompass a more thorough aspiration to dominate Ukraine politically, militarily, and economically. Ukraine for its part has strongly maintained a consolidated civic identity as a robust democracy even through the 2014 "Revolution of Dignity" and Russian annexation of Crimea, and the current war is believed to have further united Ukrainian citizens from all regions, linguistic and religious backgrounds while reinforcing the split between Ukrainian and Russian identities.

2.2 Global Public Opinion through Traditional Media

By analyzing news articles from traditional media, I were able to extract several key findings. A new Ipsos survey [**1**] finds that there is unity in concern for Ukrainians, willingness to take in refugees, and wariness of getting involved militarily, but diverging views on sanctions and military support. On average across 27 countries, 61% think it poses a significant risk to their country and to the world. Majorities in every one of the countries surveyed support taking in Ukrainian refugees (72% in the United States) and agree their country should avoid getting involved militarily in the conflict (65% in the U.S.). However, opinions on economic sanctions and providing weapons to the Ukrainian military differ widely across countries. Differences are starker when it comes to providing military support or sending troops to Ukraine. On average globally, about one-third support their country providing weapons – such as guns and anti-tank weapons – to the Ukrainian military (36%), providing funding to the Ukrainian military (33%), and sending troops to NATO countries neighboring Ukraine (32%). However, those who support sending their own troops to Ukraine are a minority in each one of the 27 countries. DG COMM's Public Opinion Monitoring Unit's survey [**2**] conducted in EU also concluded that there is continued support for Ukraine but increasing criticism towards military aid and sanctions. There is a general fear of conflict expanding to the scale of a World War. Cost of living crisis also remains a top concern and effects of inflations are increasingly reflected in consumer behavior. I hope to use our sentiment analysis to confirm or oppose these claims made by traditional media and survey results.

3 Sentiment Analysis Methodology

3.1. ML Classification Algorithms

The sentiment analysis begins with cleaning and preparing a general social media dataset, *sentiment_analysis.csv*, which contains generic tweets that have been pre-classified into negative and positive sentiments (0 and 1). A natural language toolkit (NLTK library package) in Python was used to clean the data and then fed into a text feature extraction model, TF-IDF. Performing this feature engineering/extraction provides a reasonably sized dataset with numeric variables that I can use to train binary ML classification models.

Several ML classification models (Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, Gradient Boost and Neural Networks) were trained to predict the sentiment of tweets within the TF-IDF vectorized dataset, which was split into training and testing sets for computation. These supervised ML models were all fit to the training data and implemented on the test data to assess the predictive accuracy of tweets' sentiments, results are presented in Table 1 below.

Table 1: Performance of ML algorithms on test set

Supervised Classification Model	**Accuracy (Test Dataset)**	**Implementation Time (s)**
Logistic Regression	0.9252	30.7
Naïve Bayes	0.8625	7.8
Decision Tree	0.9179	746.3
Random Forest	0.9313	1278.7
Gradient Boost	0.9164	150.5
Neural Network	0.9227	1148.7

Following the model implementation, *Random Forest* model with 100 tree estimators was selected as the best model. A grid search was used to improve the model performance through hyperparameter tuning with a 5-fold cross-validated training sets. The criterion function that measures the quality of the split and the bootstrap setting which determines if bootstrapped samples or the whole dataset should be fit to the model, were tuned. Results are presented in Table 2.

Table 2: Results of hyperparameter tuning of Random Forest model

Parameter 1 - Criterion	**Parameter 2 - Bootstrap**	**Mean Accuracy**	**Mean F1**	**Implementation Time (min)**
gini	True	0.9302	0.9480	15.44
entropy	True	0.9310	0.9482	14.66
log	True	Nan	Nan	0.70
gini	False	0.9288	0.9464	25.26
entropy	False	0.9290	0.9468	23.76
log	False	Nan	Nan	0.70

With the best model parameters for split criterion and bootstrap set to *entropy* and *True* I can ensure sufficient information that indicates the disorder of the vectorized words with the sentiments within our datasets.

3.2. Datasets Tested

3.2.1 Tweets sourced by key dates

This dataset includes sourced tweets from multiple dates since the war started [**3**]. I decided to extract tweets from milestone dates in the war, which would allow us to have diverse opinions and reactions to various events in the war, which would help us better analyze the general public's opinion on the war. The dates of tweets extracted are:

1. Feb 24, 2022: President Vladimir Putin announces a "special military operation" in eastern Ukraine.
2. Mar 25, 2022: Russia announces that the first stage of its military operation is completed, and their primary focus is now centered on "the liberation of Donbas", basically redefining their initial war aims.
3. April 6-7, 2022: The Biden administration prohibits US investments in Russia, and calls for Russia to be expelled from G20, along with UN General assembly expelling Russia from the UN Human Rights Council.
4. Jul 3-4, 2022: Russia claims to have taken Lysychansk, as it re-expands its goals to include other cities in Ukraine, marking the start of phase 3.
5. Aug 24-25, 2022: Ukraine marks its day of independence from Soviet rule, which also marks the 6-month anniversary of Russia's full-scale invasion. There was a worry that Russia would escalate attacks on that day, but President Zelensky gave a defiant speech saying that Ukraine was reborn the day Russia decided to attack.

Due to the large size of tweets from each day, I decided to extract only 60,00 tweets from each day. If that day had less than 60,00 tweets, I included all of it in our combined dataset.

3.2.2 Reddit data

Data was web-scrapped from Reddit to provide a complementary and cross-platform insight regarding the war. The 'worldnews' community on Reddit is a moderated source of updates regarding news of international relevance. They moderate the thread to only allow facts and objective reporting of events and therefore were chosen as a good complement to the free opinions expressed in the twitter dataset.

I chose the live thread "Russian Invasion of Ukraine (February 23, 2022)" and extracted all comments and sub-comments for our analysis to obtain running updates on the war.

3.2.3 Pro-Russia Tweets

The tweets from pro Russia's perspective are also important as their negative sentiments are mainly due to Ukraine's action. By analyzing it, I can better understand Russia's accusations against Ukraine. Even though they may be extreme cases, I believe this information is helpful as pro Russia's opinion will have an impact and lead to general public opinion.

To get the tweets, I extracted the tweets containing the following hashtags #IStandWithRussia, #IStandWithPutin, and #supportrussia. From the preliminary research, I saw that those supporting Russia always like to use those hashtags. After cleaning the tweets, I eliminated duplicate results, as it is common to retweet a tweet many times.

Furthermore, I noticed the hashtag #zelenskywarcriminal has the highest appearance in this dataset. Therefore, I extract a dataset with this hashtag alone from twitter as well.

3.3. Performance of pre-trained sentiment classification model (VADER) vs Random Forest

The pre-trained sentiment classification model used in our project was VADER. It is a fully open-sourced lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media [**4**].

In our analysis, as the training dataset only has two sentiment classes – positive and negative, the VADER model needs to be altered to remain consistent with the training set for better comparison between the models. Usually, VADER model outputs three classes – positive, negative, and neutral. This is tuned by a threshold value that is used to compare with the compound scores given by the VADER model. By changing the threshold value, a two-class output can be achieved.

The best model from part 1 (Random Forest) was extracted and applied to our datasets, it yielded an average accuracy of 0.56 compared to the VADER model. By looking at some of the labeled texts, it is clear that the Random Forest model is not able to classify the texts as correctly as VADER. For example, I got "Happy 4th of July", "independent day" as negative sentiments, "nuclear power plant" as positive sentiments, etc, which contradicts our logic and made it hard for further analysis. The decrease in accuracy compared to its performance in section 3.1 might be due to the nature of the training set where there were more positive texts compared to tweets and reddit comments on the topic of war (mostly dominated by negative sentiments). Therefore, I selected VADER model for sentiment analysis onwards.

4 Key Findings

4.1 Foreign Perceptions of Ukraine in the International scene

There were several relevant observations made regarding the public opinions of the international populace about Russia's invasion of Ukraine.

1. Widespread support for Ukraine: Much of the populace has a pro-Ukraine stance and is sympathetic to Ukraine's cause and plight. Keywords such as ['friend', 'support', 'thank', 'people', 'slavaukraini', 'standwithukraine'] reflect a positive attitude towards Ukraine and the way they have handled the situation. Word cloud visualizations of the hashtags from both negative and positive tweets in figure 1 reflect a high level of solidarity with Ukraine and deep condemnation and rejection of Russia's invasion and massacre of innocent civilians in Ukraine. It contains phrases like #standwithukraine, #stoprussia, #stopputin and cities that have been invaded by the Russian military force such as #mariupol, #kharkhiv, #bucha, etc.

Figure 1: Wordcloud of hashtags from Russian Ukraine war

There have also been widespread calls for and interest in NATO's involvement in aiding Ukraine, seen from the high frequency of occurrence of 'nato' in keywords and hashtags from all datasets. NATO has officially condemned Russia's brutal and unprovoked war of aggression against Ukraine, a close NATO partner, which seems to be driving positive sentiment.

2. Fear of conflict expansion is mounting: One of the prominent results from our analysis has been the growing fear of a prolonged Russia-Ukraine war and its risk to the world. Based on our topic extraction from the datasets using NMF and negative hashtag extraction below, I were able to identify keywords such as ["russiaukraine', 'russia', 'stop', 'wwiii', 'world', 'russiaukraineconflict', 'worldwar', 'putin', 'war'] alluding to concerns that people's own countries and lives are at risk. This was also observed in the n-grams with 'worldwar', 'wwiii' etc. This confirms the observations made through traditional media in Section 2.2 and provides proof that there has been a rise in nationalist attitude in the international audience as the war has raged on with increasing concern about safeguarding their way of life.

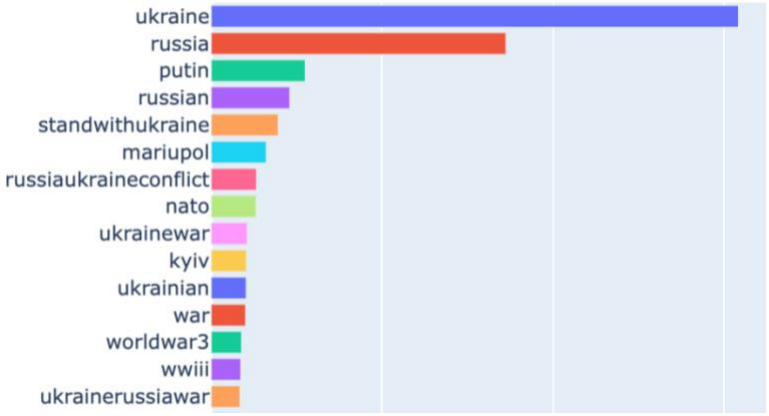


Figure 2: Common hashtags in negative tweets from Russian-Ukraine war

It would be in Ukraine's interest to highlight the impact on other countries' domestic strategic foreign policy objectives as the war can be seen as part of a battle over the future world order. Even outside of Ukraine, the war also continues to have a major effect on the global markets and food supply. Ukraine's inability to export grain throughout the first five months of the conflict worsened a global hunger crisis, with catastrophic impacts throughout the world. Russia's war of aggression against Ukraine has also triggered the biggest energy price shock since the 1970s, which is weighing heavily on the world economy. These provide justification for the need for other countries to put their domestic interests aside and offer aid.

3. Reduced engagement over time: One of the issues with Ukraine's current portfolio of strategies that can be observed is reduced rate of engagement over time on the topic of Russia's war on Ukraine. From our Twitter dataset time-series, I can observe an overall reduction in the total tweets, be it negative or positive. Also, although most tweets remain positive, they have reduced by more than 50% between May and August 2022. This can prove to be detrimental to Ukraine's cause in the long-term.

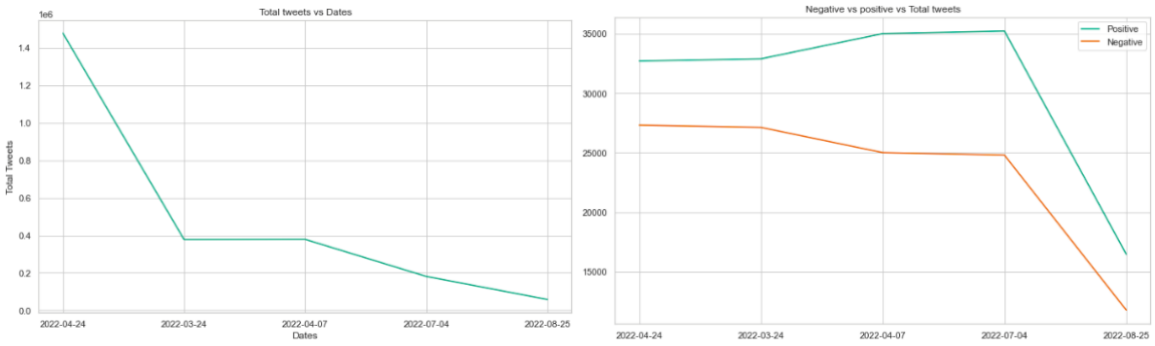


Figure 3: Timeseries of a) Overall tweets and b) Labelled tweets

Given the worsening crisis in Ukraine, more humanitarian assistance is needed to combat it. Since the war began, access has been extremely challenging in areas beyond the control of the government of Ukraine and Ukraine needs foreign parties to facilitate humanitarian response to make sure every person who desperately needs assistance is supported. Therefore, maintaining public interest is paramount to ensuring continued international aid. Keeping the international media informed and engaged with interviews and updates for strategic communications would be the best way to achieve this. President Zelensky's communication with political leaders and general audience at the start of the war has been lauded for its effectiveness and should be continued to drive engagement. President Zelensky is a prime example of how a modern leader can rally others to his side through a simple message of humanity and shared mission and savvy in communicating that to the world, and Ukraine should utilize this advantage to its fullest.

4.2. Sentiment driver- Nazi Ukraine

One of the topics that I noticed was mentioned in the pro-Russian tweets' dataset, which is a topic related to a comparison between Ukraine and the Nazis. The 2-word phrase N-gram for negative sentiment tweets in Figure 4 shows phrases like "istandwithrussia naziukrain", "zelensky war criminal" and "nazis ukraine". While also the topic extraction of negative sentiment showcasing words like "nazis", and "asuvnazis". President Putin gave a lot of reasons trying to justify his invasion of Ukraine, and one of the reasons that raised a lot of attention and confusion is that Ukraine is run by neo-Nazis or what I can call far-right groups. Putin basically claimed that those far-right groups are secretly running the Ukrainian government and committing crimes against Russian speakers in Ukraine. Part of his incentives to invade Ukraine was for the "demilitarization and denazification of Ukraine" and to "protect people" who have been subjected to genocide. A lot of Russian state media and people on social media were promoting this propaganda, as I also noticed from our analysis on the pro-Russian dataset.

The purpose of all this propaganda from the Kremlin is to manipulate the international public opinion of Ukraine, vilify and demonize it in the eyes of the world. This is all far from the truth. Based on a post written by the US Department of state [**5**], 140 international historians have denounced this claim by the Russian government, calling this narrative factually wrong and very offensive to the victims of Nazism. Moreover, President Zelensky is Jewish, with family members that were killed by Nazis, and the pro-Russians were still trying to delegitimize his Jewishness and tarnish the image of the Ukrainian government to justify the invasion.

So, to spread the truth and falsify these claims made by the pro-Russian social media users and the Russian government, media outlets and the Ukrainian government representatives need to educate people about the Ukrainian government, its history and the people running it. Social media users should be aware that the people of Ukraine democratically elected a Jewish president about three years ago, which does not make sense that the people electing a Jewish president would have any ties with a Nazi (far right) group. Obviously, there is far right groups in every country including Ukraine, but they are very marginal and do not have any political influence or a say in the way the government is running the country. Scholars who spend their careers studying Nazism, and remembering their victims are strongly denying these allegations towards the Ukrainian government. These type

of studies and articles should be shared more often to educate people about the truth and deny all the propaganda being spread by the pro-Russian social media users.

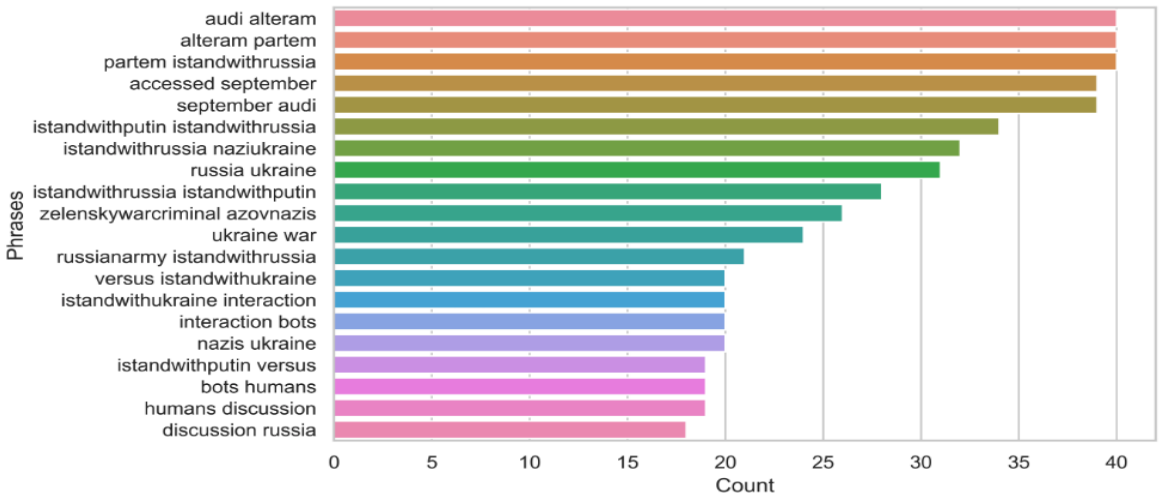


Figure 4: 2-word N-gram for negative sentiment tweets

4.3 Sentiment driver – Propaganda against international aid

One keyword appearing most frequently from the support Russia tweets starting from Feb 2022 is zelenskywarcriminal. I were interested in how the pro Russia thinks that way, and further extract the dataset using hashtag #zelenskywarcriminal.

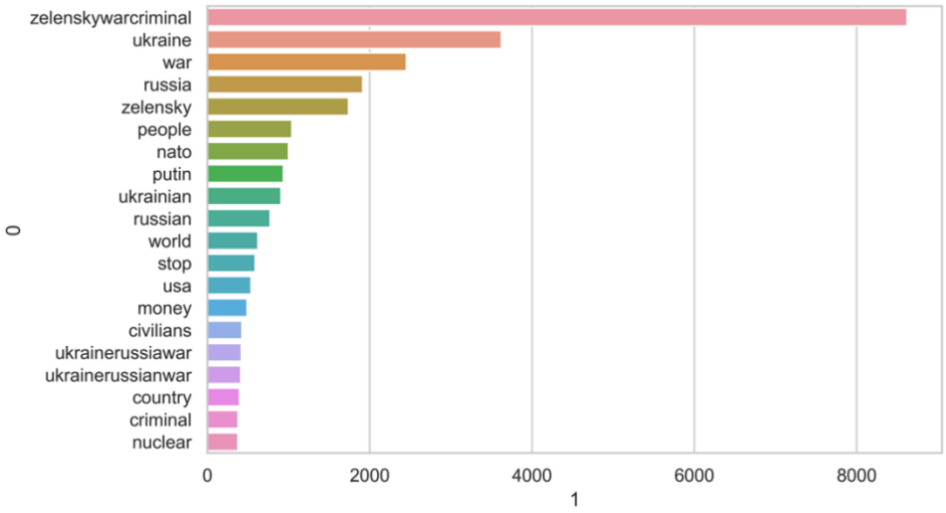


Figure 5: 1-word N-gram of #zelenskywarcriminal hashtag

Within those keywords, I think that "stop", "money", "usa" might link to money-related events. I further went to look at the keyword "money" using a word cloud within the #zelenskywarcriminal dataset.

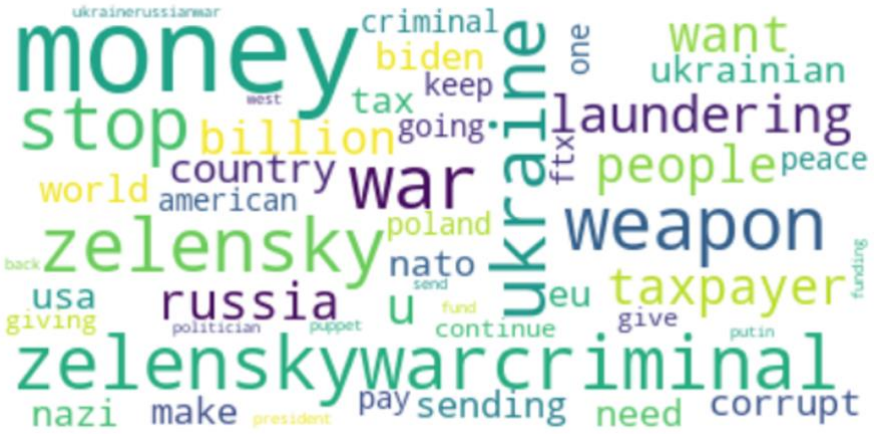


Figure 6: Wordcloud of #zelenskywarcriminal tweets with keyword money

From the word cloud, I also saw some interesting words such as "FTX", "laundering", "money", "taxpayer", "corrupt". I then analyzed the news and found out those keywords are linked to recent FTX events. A cryptocurrency exchange called FTX recently filed for bankruptcy. Since Ukraine is using crypto as a donation pathway, some people wrongly linked Ukraine with different conspiracy theories.

One conspiracy is that Ukraine is using FTX to do money laundering for the US. According to the AP News Verification website, the statement is "So Biden gave loads of money to Ukraine, who gave loads of money to FTX, who gave loads of money to Democrats." [**6**] While this statement is incorrect and misleading, people still use it as an excuse to frame Ukraine and further damage its image. The fact is that the CEO of FTX, Sam Bankman-Fried, donated money to both Democrats and Republicans [**7**]. Also, Ukraine has not stored or invested in FTX but used it to convert the donation to local currency, according to its Ministry of Digital Transformation.

A major propaganda people use to accuse the Ukraine government is a lack of transparency on the use of the donation. People came up with a conspiracy that the money was put back to invest in FTX and help the US launder money. With over 50 million USD worth of crypto donated to the Ukraine government [**8**], it is important to show how the government spends the contribution. One way is to provide a detailed description on the transaction hash so that it can be traceable. At this moment, it is hard to locate what the crypto was spent on by reading the transaction on etherscan.io because it lacks details. The information available on the internet is a picture posted by @FedorovMykhailo, a Ukraine government official, saying that the money was used to buy a certain amount of equipment [**9**], without further proof.

Blockchain has the advantage of being transparent, and the government should take advantage of that so that the public knows where the money goes. One way to do this is to attach a receipt or image of the purchase without disclosing any military information perhaps off-chain, and a brief transaction description of where the withdrawal has been used on. When being fully transparent, the conspiracy, such as Ukraine having a connection with FTX, will be minimized.

Finally, to prevent catastrophic damage that FTX bankruptcy bought to the citizen and the government, it is best to store the crypto on a cold wallet instead of centralized exchanges like Kuna Exchange. No matter how promising the exchange is or how many assets they are backed by, it can still be vulnerable to attack and policy changes. Therefore, I encourage the Ukraine government to use a cold wallet to store its cryptocurrency and promote its importance to its citizens.

Blockchain is immutable, permissionless, and fully traceable. With the proper use of it, money-related conflict and confusion can be solved easily. However, it is up to the Ukraine government to decide on how much information to put on the blockchain and how easy it is for the public to access. Moreover, whether Ukraine will

store their crypto into a cold wallet and educate their citizen on the important of cryptocurrency safety. With the emerging blockchain technology, proper use of it can bring a positive image rather than a false accusation.

5 Recommendations

Several suggestions to the Ukrainian defense strategies were explored that are projected to positively impact the Ukrainian international image and rapport. Analyses with our trained supervised ML algorithms revealed details to both positive and negative sentiments of stakeholders to the war and these same tools could be used to reaffirm public opinion and reassure those perceiving the news. Our recommendations to the Ukrainian government are highlighted below based on our key findings.

1. Active monitoring of public opinions, swift denouncing of inflammatory claims and active engagement of international media

   A data-defense strategy involving sentiment analysis could help Ukrainian officials. and denounce inflammatory claims made to fuel escalating tensions. Keeping the international media informed and engaged with interviews and updates for strategic communications will also maintain public interest and eliminate misconceptions

2. Sensitization against the Nazi-Ukraine propaganda

   Our sentiment analysis also highlights the importance of education in the Ukrainian defense strategy. Our models detected several sentiments portraying Nazi-Ukraine propaganda as one of the instigations for the conflict, emphasizing the need for media outlets and the Ukrainian government representatives to sensitize the public about the Ukrainian government, its history and the people running it. Relevant official articles archived from important time periods need to be preserved and shared to educate people about the truth and deny all the propaganda being spread by the pro-Russian social media users.

3. Use of foreign aid should be made more transparent by utilizing the power of financial technologies

   Lastly, sentiment analysis highlights the false claims and conspiracies about Ukraine connection with FTX, money laundering for the US and undisclosed use of foreign aids. Tools like new financial technologies such as blockchain, could help maintain global transparency in Ukraine's use of foreign aid. Since Ukraine only uses the FTX to convert foreign donations into her local currency, it's important to provide some description of the transaction hash of the blockchain so that the public can view and know how the donations have been utilized. Furthermore, to prevent catastrophic damage that FTX bankruptcy brought to the citizen and the government, it is best to store the crypto on a cold wallet instead of centralized exchanges like Kuna Exchange.

6 References

[1] Ipsos, "Global Public Opinion about the war in Ukraine", April 19, 2022, https://www.ipsos.com/en-us/news-polls/war-in-ukraine-april-2022, Date assessed: December 2, 2022.

[2] DG Communications Public Opinion Monitoring Unit, "Public Opinion on the war in Ukraine", November 24, 2022, https://www.europarl.europa.eu/at-your-service/files/be-heard/eurobarometer/2022/public-opinion-on-the-war-in-ukraine/en-public-opinion-on-the-war-in-ukraine-20221124.pdf. Date assessed: December 2, 2022

[3] Aljazeera News Agency, “Russian-Ukraine war - List of key events, day 279, November 29, 2022, ”https://www.aljazeera.com/news/2022/11/29/russia-ukraine-war-list-of-key-events-day-279, Date assessed: December 1, 2022

[4] Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014

[5] US Department of State, “To Vilify Ukraine, The Kremlin Resorts to Antisemitism, July 11, 2022, ”https://www.state.gov/disarming-disinformation/to-vilify-ukraine-the-kremlin-resorts-to-antisemitism/, Date assessed November 30, 2022

[6] Karena Phan and Ali Swenson, “Post spin baseless theory about FTX, Ukraine and Democrats”, AP NEWS, November 17, 2022, https://apnews.com/article/fact-check-ftx-ukraine-democrats-laundering-996663874380?fbclid=IwAR2SeIXjYTEOYc5onZZLu8uJjvDTLNjjBYtKDozwcu_kNgligBBqJ3cD5VI, Date assessed: December 1, 202

[7] Marco Quiroz-Gutierrez, “Sam Bankman-Fried says he donated just as many millions to Republicans as Democrats, but didn’t publicize it because reporters would ‘freak the f–k out”, Fortune Crypto, November 29, 2022, https://fortune.com/crypto/2022/11/29/sam-bankman-fried-political-donations-democrats-republicans-dark-money/, Date assessed: November 30, 2022

[8] Brayden Lindrea, “Aid for Ukraine’s $54M crypto fund buys vests, scopes and UAVs”, Coin Telegraph-The Future of money, August 18, 2022, https://cointelegraph.com/news/aid-for-ukraine-s-54m-crypto-fund-buys-vests-scopes-and-uavs, Date assessed: December 1, 2022

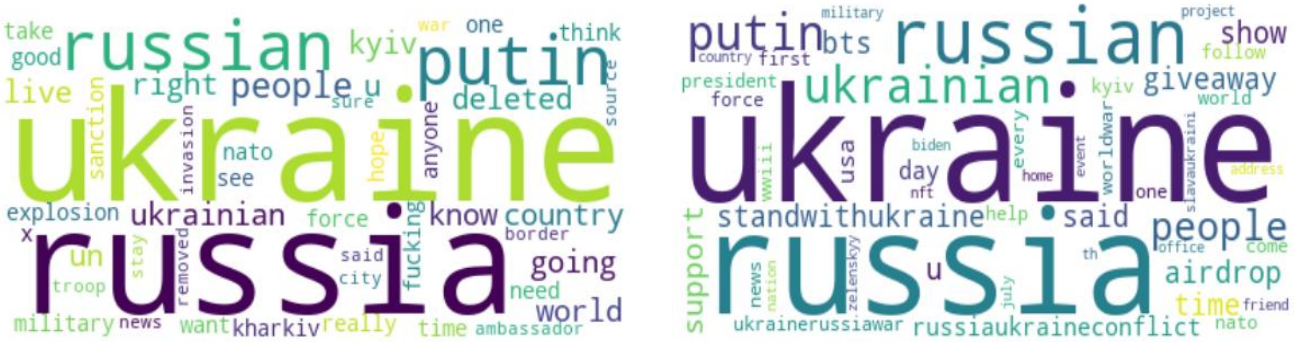
[9] Mykhailo Fedorov-Ukraine Government Official, “Crypto Fund aid for Ukraine total cost”, Twitter, August 17, 2022, https://twitter.com/FedorovMykhailo/status/1559923193581899777/photo/1, Date assessed: December 2, 2022

7 Appendix

Visualizations

1. Wordclouds

Figure 1. a) Vader model wordclouds on all positive tweets (left); b) on all negative tweets (right)



c) Vader model wordclouds on all positive reddit comments (left); d) on all negative comments (right)

Figure 2. a) Vader wordcloud on zelenskywarcriminal dataset

2. N-grams

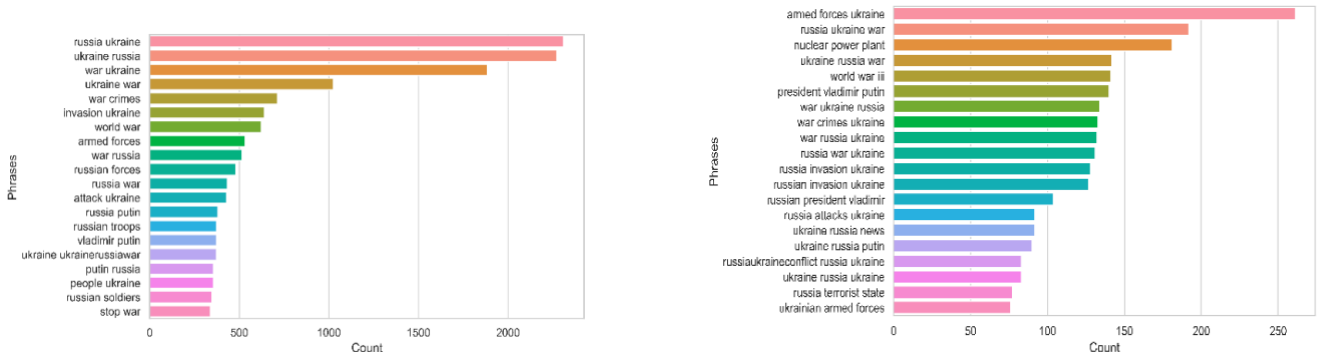


Figure 1. a) Vader model 2-gram on all negative texts (left); b) 3-gram on all negative texts (right)

Figure 2. a) Vader model 2-gram on all positive texts (left); b) 3-gram on all positive texts (right)

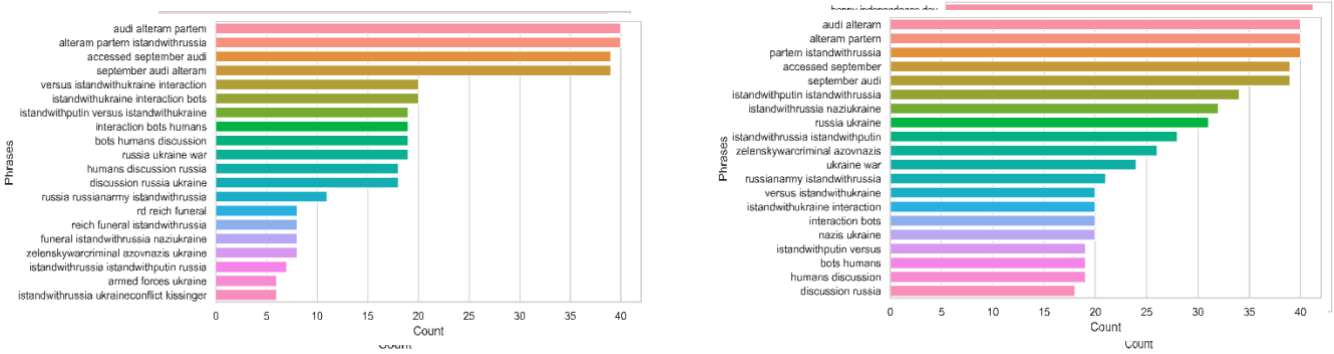


Figure 3. a) Vader 2-gram on negative pro-Russia texts (left); b) 3-gram on negative pro-Russia texts (right)

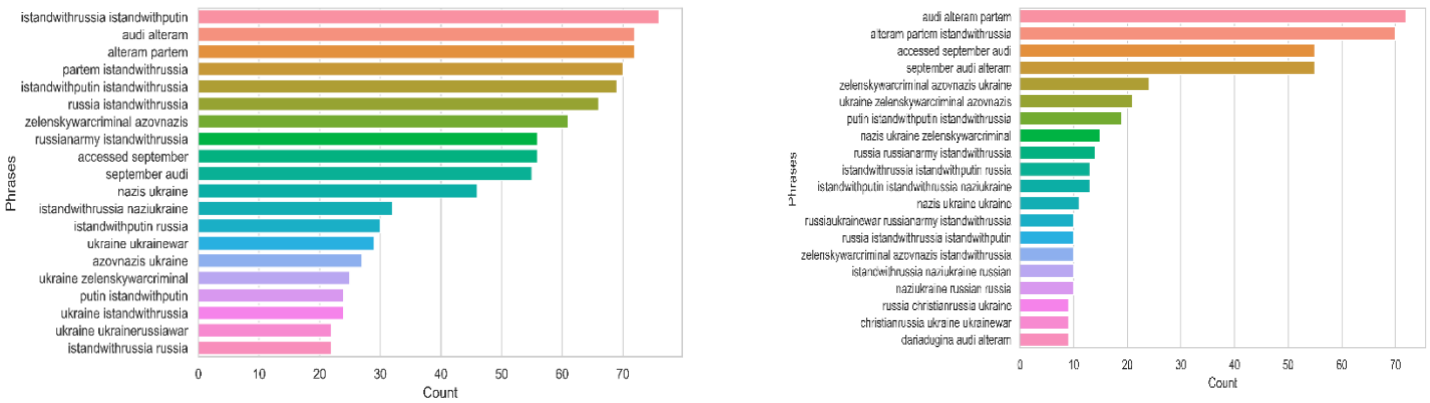


Figure 4. a) Vader 2-gram on positive pro-Russia texts (left); b) 3-gram on positive pro-Russia texts (right)

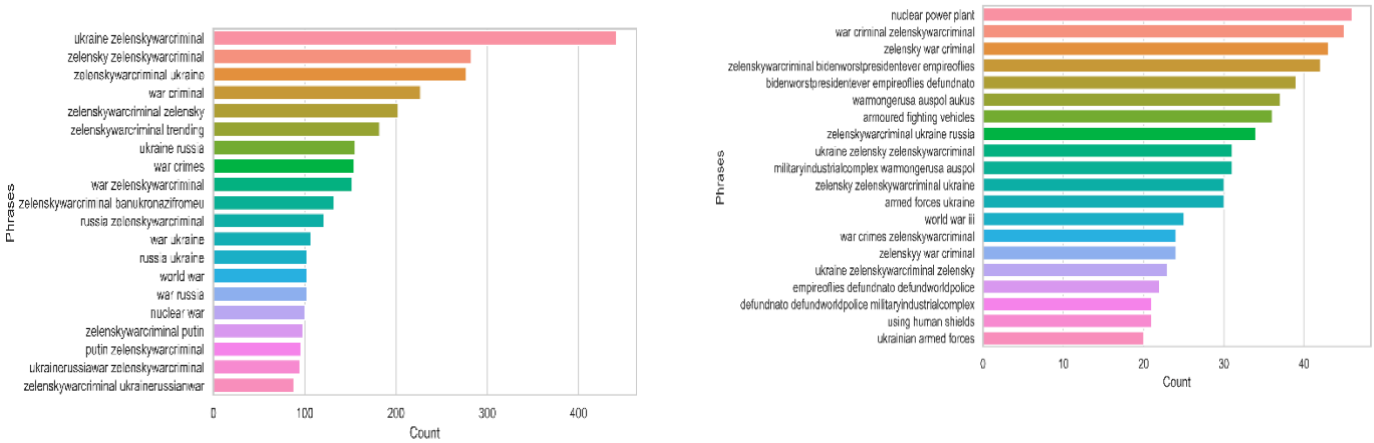


Figure 5. a) Vader 2-gram on zelenskywarcriminal texts (left); b) 3-gram on zelenskywarcriminal texts (right)

3. Barplots

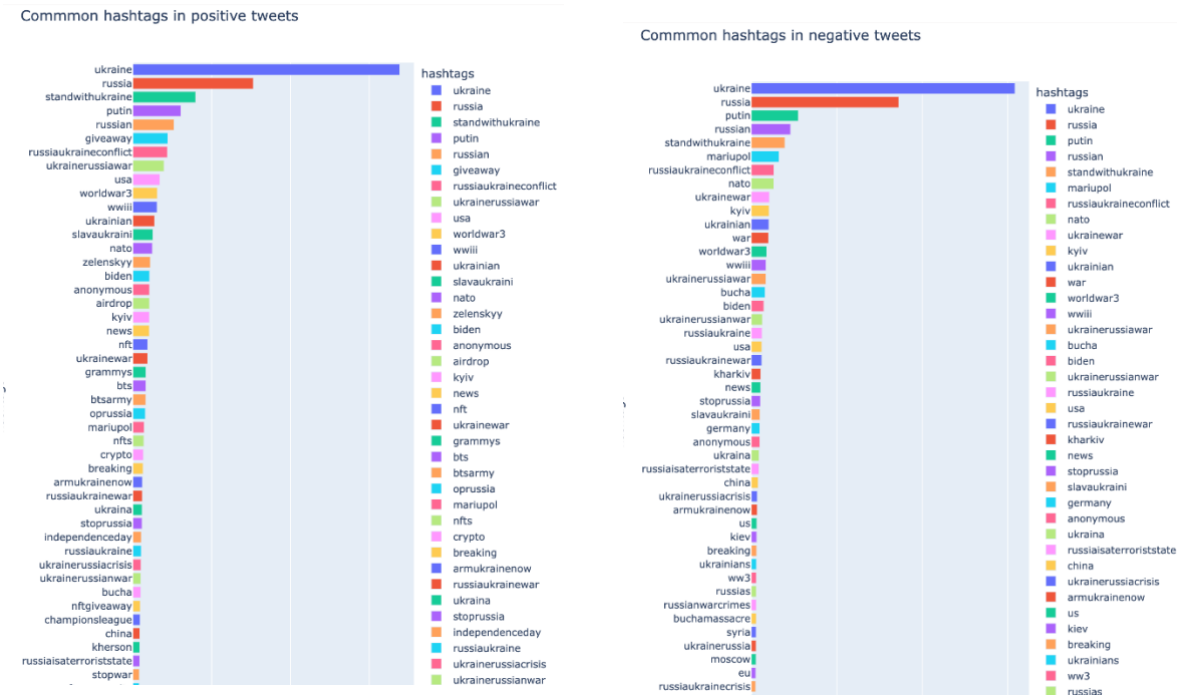


Figure 1. a) Barplot of popular hashtags in positive tweets b) in negative tweets