Public Sentiments on Russia-Ukraine Conflict: Sentiment Analysis of Twitter Posts (1640 words)

Akshat Arvind

Indiana University

1. Introduction

The tension between the two nations – Russia and Ukraine started to rise in the starting months of 2022. In the month of February, when Russia started invading Ukraine, it invited lot of reactions from people across the world on social media especially on Twitter. The hashtags #UkraineRussiaWar, #Ukraine, #RussiaUkraineConflict has been on constant trending for the month of February and March 2022. Twitter as a social media platform hosts a lot of data in the form of tweets from public opinion on such incidents.

This work aims to perform a sentiment analysis on the collected tweets related to the Russia Ukraine war to better understand the public sentiment on a serious war situation. Previous research work has been done in this area including an analysis on tweets focusing on sentiment analysis on the Syrian Refugee crisis (*Ladner, Karsten & Ramineni, Ruchishya & George, K.M. (2019). Activeness of Syrian refugee crisis: an analysis of tweets. Social Network Analysis and Mining*) and analyzing public opinion on refugee and humanitarian policy (*Blitz, B. (2017) 'Another Story: What Public Opinion Data Tell Us about Refugee and Humanitarian Policy', Journal on Migration and Human Security*).

2. Research Questions

The tweets collected and analyzed were used to answer the below research questions: -

i) As the conflict between the two nations progressed, how did the overall sentiment for one nation compared to another?
ii) Did the sentiments for both the nations change from February 2022 to March 2022?

3. Data

The required data for this research was collected from Twitter in the form of tweets. Total 2000 tweets were collected from the months of February and March, and the collection criteria was set as hashtags *Ukraine, Russia, War, Putin, Zelensky*. The tweets were collected using a python program using the 'snscrape' python package. Since the analysis demanded access to historical tweets, ‘snscrape’ package proved to a better choice to avoid twitter’s rate-limiting and time period limitations on their Twitter API.

Tweets selection criteria:

1. The objective of this study is to analyze the change in people’s sentiments on the Russia-Ukraine conflict as the conflict progressed. So, the tweets from early February 2022 to end of March 2022 satisfying the query keyword criteria has been included in the study.

2. The tweets that contain keywords (Russia, War, Putin) were used for Russia and keywords (Ukraine, War, Zelensky) were used for Ukraine sentiment analysis.

3. To avoid the bias in the data, when searching for keywords related to ‘Russia’, keywords related to ‘Ukraine’ were excluded and vice-versa.

4. 1000 tweets for each keyword were collected from months of February and March and stored in a CSV file for further analysis.

4. Methods

For this research, sentiment analysis was conducted using VADER (Valence aware Dictionary and Sentiment Reasoner) tool available as a python module in the python library. VADER is a lexicon and rule-based sentiment analysis tool that is designed for analyzing sentiments in social media. VADER analyzes the sentiment of the text and labels the post as positive, negative, or neutral. It does so by assigning a compound score which ranges from -1 to +1. These scores help in classifying the posts into categories like very positive, positive, neutral, negative, and very negative.

Reason for choosing VADER as the tool for sentiment analysis is its popularity and accuracy with which it can analyze sentiments. The original research paper (*Hutto, C.; Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. ICWSM 2014, 8, 216-225.*) describes VADER, as a simple rule-based model for general sentiment analysis. VADER’s effectiveness was compared with eleven state-of-practice benchmarks including LIWC, ANEW, SentiWordNet, and ML oriented techniques relying on Naïve Bayes, SVM algorithms. VADER outperformed individual human raters (F1 Classification Accuracy = 0.96 and 0.84, respectively) and generalizes more favorably across contexts than any of the other benchmarks.

The Sentiment analysis study was conducted through below steps:

a) Creating Primary DataFrame from downloaded CSV File:

Pandas Python library was used to read the CSV file and create a dataframe for filtering, grouping and aggregate operations. The 'tweet_date' column was converted into a datetime object to extract 'tweet_month' which was added as a new column to help in the analysis.

Keywords	Text	tweet_date	tweet_month
Russia	@SamRamani2 How much does Russia earn daily on...	2022-03-28	3.0
Russia	Listen @NPR @JennaMC_Laugh interview w/Belarus...	2022-03-28	3.0
Russia	@InnaSovsun I fully understand you, because mo...	2022-03-28	3.0
Russia	@Reuters What wrong USA have large oil, gas ! ...	2022-03-28	3.0

Sample of DataFrame with tweet_month

b) Determine the sentiment and add a label for each data point in the dataset:

Each tweet in the dataset was analyzed by the SentimentIntensityAnalyzer method available under vader.vaderSentiment module. SentimentIntensityAnalyzer returns a compound sentiment score between -1 to 1 for each tweet, where -1 denotes very negative sentiment and 1 denotes positive sentiment. The compound scores generated were used to create five categories to classify the tweets into. The created 5 categories and their criteria is shown below:

Greater than 0.5:	Very positive
Greater than 0 and less than or equal to 0.5:	Positive
Equal to zero:	Neutral
Less than zero and greater than or equal to -0.5:	Negative
Less than -0.5:	Very negative

Russia	Latvia’s deputy prime minister tells it like i...	2022-03-28	3.0	very negative
Russia	@naman00099 @mfa_russia You are right that Put...	2022-03-28	3.0	positive
Russia	@MikeGibbons84 @piersmorgan Amen. Piers is out...	2022-03-28	3.0	very negative
Russia	@AmirAdilkhan @humsterua @aluxnder @YourAnonNe...	2022-03-28	3.0	very negative
Russia	@mfa_russia @AviaHistory @mod_russia @RusEmbUS...	2022-03-28	3.0	negative

Sample of DataFrame with sentiment labelling

c) Aggregating the compound sentiment scores by month and Keyword:

To further continue the analysis, the sentiment scores were grouped based on month and keyword to derive insights from the sentiment analysis. The number of tweets were grouped based on the keyword and month and aggregated sentiment was calculated for each criterion. One such example for keyword ‘Ukraine’ for the month of February is shown below:

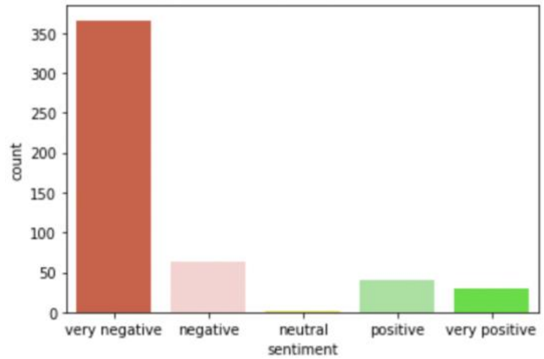
```
very negative    219
very positive     98
negative          80
positive          56
neutral           48

very negative    43.7%
very positive    19.6%
negative         16.0%
positive         11.2%
neutral           9.6%
```

d) Plotting and presenting the results:

To present the results in this paper, Python data visualization library ‘Seaborn’ has been used. It is a simple tool which allows making elegant charts with multiple customizable options. The ‘countplot’ method in the Seaborn library has been used to display results in this paper.

The order parameter was used to predefine the order of sentiments to be displayed on the charts i.e. (*very negative, negative, neutral, positive, very positive*). Also, the palette parameter was used to define color scheme to easily distinguish between the positive and negative sentiments.

One such example for keyword 'Russia' and the month of 'March' is shown below:

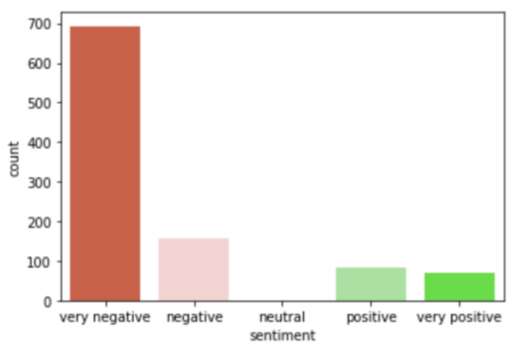


Count plot showing sentiment distribution for Russia in the month of March

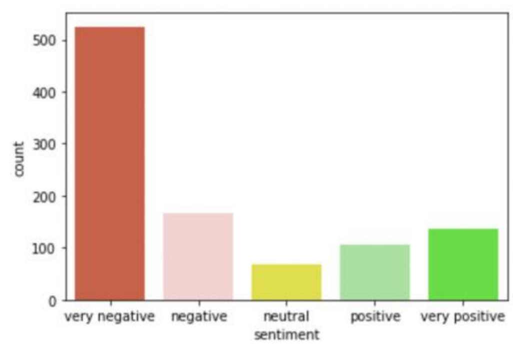
5. Results

Analyzing overall sentiments on Russia and Ukraine

The charts below show the overall sentiment distribution related to keyword 'Russia' and 'Ukraine'.



Russia



Ukraine

The data below shows the sentiment distribution for the tweets containing the keyword 'Russia' and 'Ukraine'

very negative	69.2%
negative	15.5%
positive	8.3%
very positive	7.0%
neutral	0.1%

Russia

very negative	43.7%
very positive	19.6%
negative	16.0%
positive	11.2%
neutral	9.6%

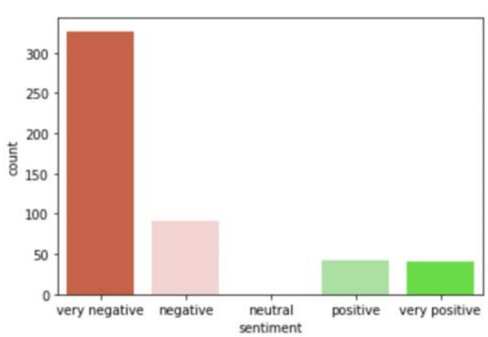
Ukraine

Discussion:

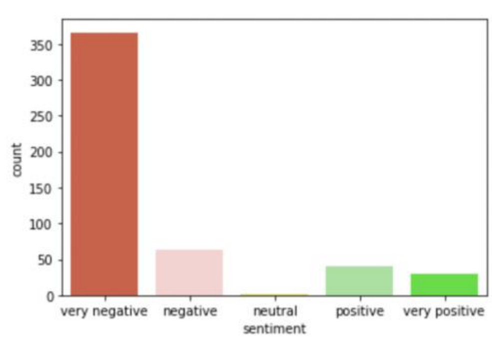
After analyzing the overall sentiments for the tweets on Russia and Ukraine, the results indicate that the overall negative sentiments related to tweets about Russia comprise about 85% of the total tweets, whereas for Ukraine the overall negative sentiments distribution is around 64%.

Analyzing the change in sentiments for Russia from February to March

The charts below show the sentiment distribution related to Keyword ‘Russia’ for the month of February and March:



Russia – February



Russia – March

The data below shows the sentiment distribution for the tweets containing the keyword ‘Russia’ for the month of February and March:

very negative	65.3%
negative	18.4%
positive	8.4%
very positive	8.0%

Russia – Feb

very negative	73.1%
negative	12.6%
positive	8.2%
very positive	6.0%
neutral	0.2%

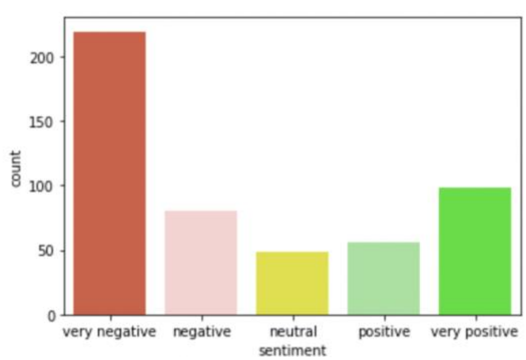
Russia – Mar

Discussion:

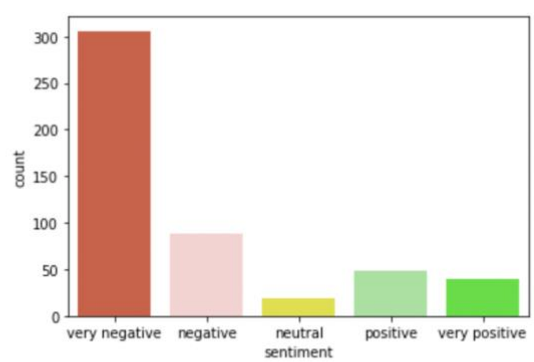
After analyzing the results for the sentiments distribution for the tweets containing the keyword ‘Russia’, the results indicate that tweets containing negative sentiments increased silightly from 83% from Feb to 85% in March, but the very negative tweets percentage increased from 65% in Feb to 73% in March.

Analyzing the change in sentiments for Ukraine from February to March

The charts below show the sentiment distribution related to Keyword ‘Ukraine’ for the months of February and March:



Ukraine – February



Ukraine – March

The data below shows the sentiment distribution for the tweets containing the keyword ‘Ukraine’ for the months of February and March:

very negative	43.7%
very positive	19.6%
negative	16.0%
positive	11.2%
neutral	9.6%

Ukraine – Feb

very negative	61.1%
negative	17.6%
positive	9.8%
very positive	7.8%
neutral	3.8%

Ukraine – Mar

Discussion:

After analyzing the results for the sentiments distribution for the tweets containing the keyword ‘Ukraine’, the results indicate that tweets containing positive sentiments decreased from 30 % from February to 17 % in March, and the negative sentiments rose from 60 % to 79 %.

6. Limitations

Though careful measures were taken to improve the quality of the data being used for sentiment analysis, there could be possibility of noise in the data such as spams like people using trending hashtags to share their own posts which is irrelevant to the hashtags being used.

Also, as the conflict has been going on for more than 2 months, and many events have happened in the timeline. There is a possibility that the selected keywords for this study might not be suffice and have missed out some important tweets. A comprehensive search for the relevant hashtags and using them could have helped in improving the quality of the data even more.

7. Conclusion

In this research, sentiment analysis was conducted on 2000 tweets collected from Twitter from Feb 2022 to Mar 2022 to study the public sentiments on Russia and Ukraine and how the sentiments changed as the conflict progressed from February to March.

Comparing the overall sentiments between tweets related to Russia and Ukraine, the results showed that tweets related to Russia had more negative sentiments attached compared to Ukraine. As the conflict progressed, the results show that the negative sentiments for Russia increased from February to March which could be expected as the news of bombings of various cities in Ukraine by Russia started to come in early March.

A surprising result from the study was the increase of negative sentiments for Ukraine from February to March. There does not seem to be an explanation for this trend. Manual inspection of the tweets might help find the reason for this trend.

After this study, it can be stated that

- Majority of the negative sentiments were related to the tweets which contained the keywords related to 'Russia'.
- As the events of the conflict progressed, the negative sentiments towards Russia increased.

8. References

Ladner, Karsten & Ramineni, Ruchishya & George, K.M. (2019). Activeness of Syrian refugee crisis: an analysis of tweets. Social Network Analysis and Mining

Blitz, B. (2017) 'Another Story: What Public Opinion Data Tell Us about Refugee and Humanitarian Policy', Journal on Migration and Human Security

Hutto, C.J. & Gilbert, Eric. (2015). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of social media Text. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014.