MIE 1624 Final Project:
Sentiment Analysis on Ukraine Russia Conflict

Group 13
Prepared by:

Xiaotong Liu

Jialin Shang

Jiekai Yin

Ruichen Lu

Mingxuan (Janet) Cui

Zhuocheng Yang

Ziwei Jia

Table of Contents

Executive Summary

In February 2022, Russian President Vladimir Putin mobilized Russian troops to invade Ukraine on the fabricated grounds of "demilitarization and denazification", which officially turned the conflict between the two countries into a full-scale war that quickly developed into the most significant war in Europe since World War II. This military action is widely seen as aggression, and the topic has been highly discussed on Twitter, a social media platform with over 200 million daily active users. We will analyze the discussion on Twitter about the war in Russia and Ukraine to provide advice on armed conflict, anti-war propaganda, people's security, and the international situation. In this project, we use the "Sentiment Analysis.csv" file provided by Prof. Oleksandr Romanko to train our sentiment analysis models and apply them to relevant tweet datasets, including "Opinion of influencers - tweets of Elon Musk and responses to his tweets" by Prof. Oleksandr Romanko and "Russia-Ukraine Conflict Twitter Dataset" by Muhammad Tariq. The selected data set shows that Internet users are mainly against the war, and most of them condemn the Russian invasion as inhuman and irresponsible. On the other hand, Internet users' concern about the war reflects their desire for peace and sympathy for the Ukrainian people. At the same time, many people strongly support the Ukrainian Patriotic War and use the words "Glory to our defenders!" and other slogans. But in general, pacifism and international aid are the main themes of Internet users' statements.

Introduction

The relations between Russia and Ukraine have become more intense since 2014, and the conflict turned into a full-scale war in February 2022. Given the ongoing war, this report aims to analyze the public opinions on the conflict from social media. In particular, we implemented sentiment analysis on various data primarily collected from Twitter. (Tariq, 2022) Considering the impact influencers can have on public opinions, we also included the tweets of Elon Musk, a celebrity billionaires and public figure, for the analysis. (Romando, 2022)

The sentiment analysis involves three major procedures: using natural language processing methods to process text data, classifying tweets into positive and negative sentiments with a machine learning model, and visualizing key information, thereby providing insights.

Problem Definition

Recently, individuals, organizations, and governments have widely employed sentiment analysis to detect public opinion trends. By analyzing social media content, managers will know factors and topics that drive sentiment, which may positively impact the end of the war. In our project, we will focus on the social media posts/tweets related to the war and find the factors, reasons and topics that drive public opinion about the war. And provide strategies to the Ukrainian government and international NGOs.

Methodology

1. Sentiment Modeling

We used "Sentiment Analysis.csv" in the first phase to train our sentiment analysis model. (Romanko, 2022) The dataset contains 550,391 tweets with their sentiments labeled. We cleaned the text by removing punctuations, links, stopwords, and retrieving normal forms of the words using lemmatization. Then we use the TF-IDF (term frequency-inverse document frequency) method to determine how relevant the words are to sentiment analysis. Initially, we selected the top 75 features (words) to classify the tweets using six models: Random Forest, Decision Trees, XGBoost, k-NN,

Logistic Regression, and Naive Bayes. Based on the confusion matrices and computational time, we selected the best four models (Random Forest, Decision Trees, XGBoost, and Logistic Regression) and performed hyperparameter tuning and cross-validation. The tuning results showed that the XGBoost classifier performs best, with a prediction accuracy of 83.84% on the validation set. The XGBoost classifier with the same tuned hyperparameters is used to classify the same dataset with different numbers of features. The purpose is to maintain accuracy > 90% on the validation set and avoid sacrificing too much computational time. The optimal model is the XGBoost classifier with max depth of 15 and feature number of 250.

2. Sentiment Classification

With the best model selected and trained in the previous part. We fitted the specific XGBoost model to the two mentioned datasets of tweet messages about the Russia-Ukraine Conflict and predicted the sentiment of each tweet message.

3. Factor and Topics Identification

The factors and reasons that determine positive and negative sentiments are identified by extracting keywords from each tweet in the dataset. The methodology used for information retrieval is TF-IDF, which reflects the importance of the words to each tweet in the whole dataset. The TF-IDF score increases as the word appears more frequently in the tweet, which is offset by the number of tweets in the dataset containing the word. The keywords with the top 3 highest TF-IDF scores are selected for each tweet, and the 50 words that appear most frequently in the selected keywords are used for text summarization.

Findings & Recommendations

1. Findings in tweets of Elon Musk Dataset

On October 3, Tesla and SpaceX CEO released a public opinion survey to his 119 million fans on Twitter, voting for his peace plan, which includes four parts, to end the war between Russia and Ukraine.

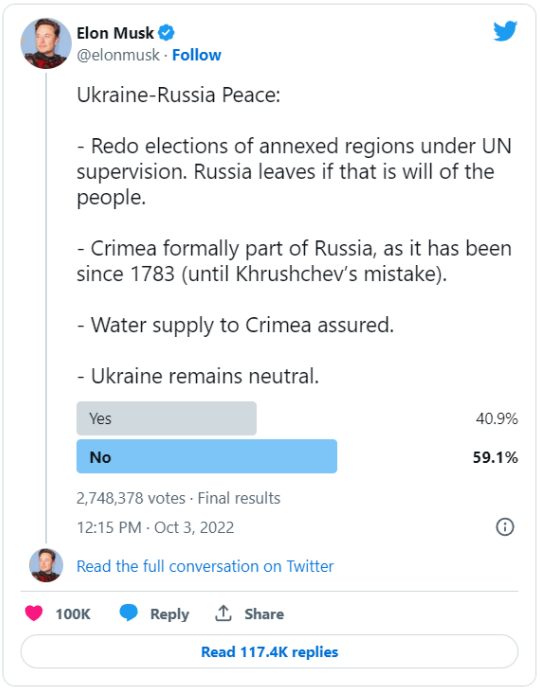

Figure 1. Screenshot of Elon Musk's Twitter Post

Our data set is to record the replies from netizens to Musk within three days after this incident, and we have also researched this data set.

The two histograms below show the most common 50 words for positive and negative messages. For the positive words, 'nan' is the most common word, with over 17500, nearly three times the second most common word, 'elon'. And for the negative messages, 'geoscale' appears almost 1400 times, approximately 1.3 times more than the second most common positive word, 'bub'. Both histograms are right-skewed, showing a smoothly decreasing trend, indicating that the rest words have a similar appearance.

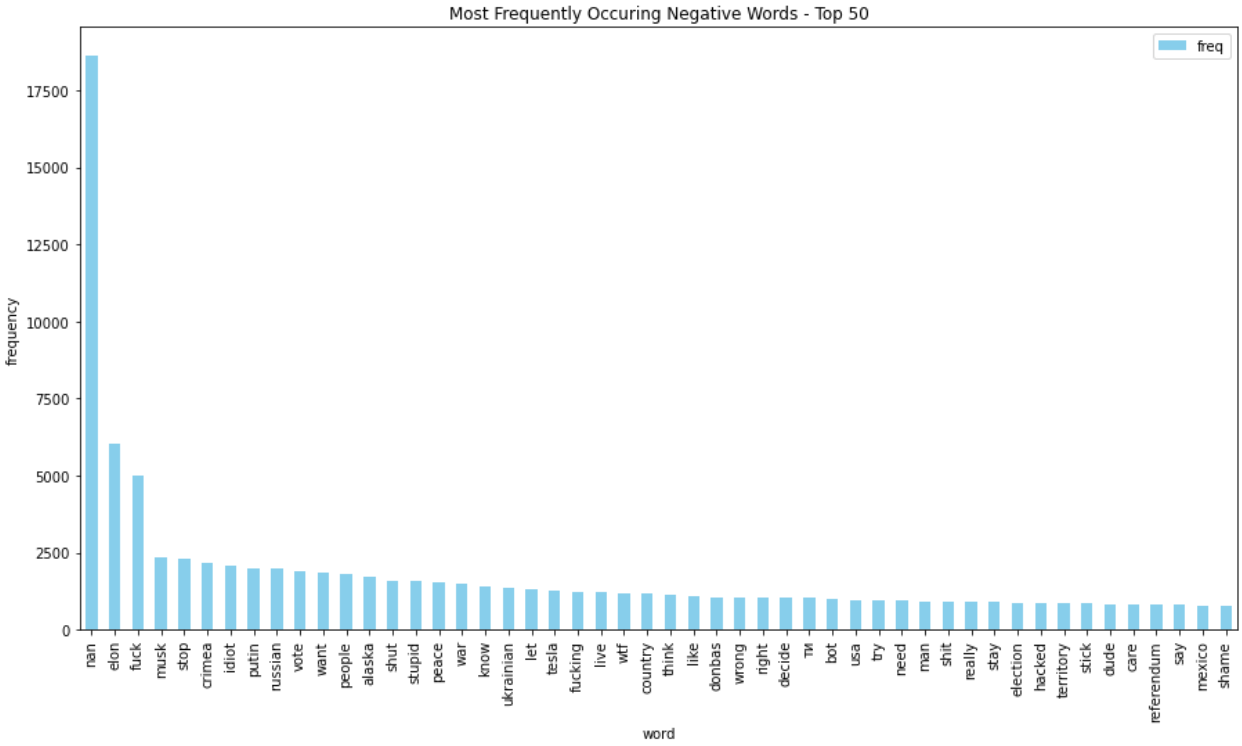

Figure 2. Top 50 Negative Words

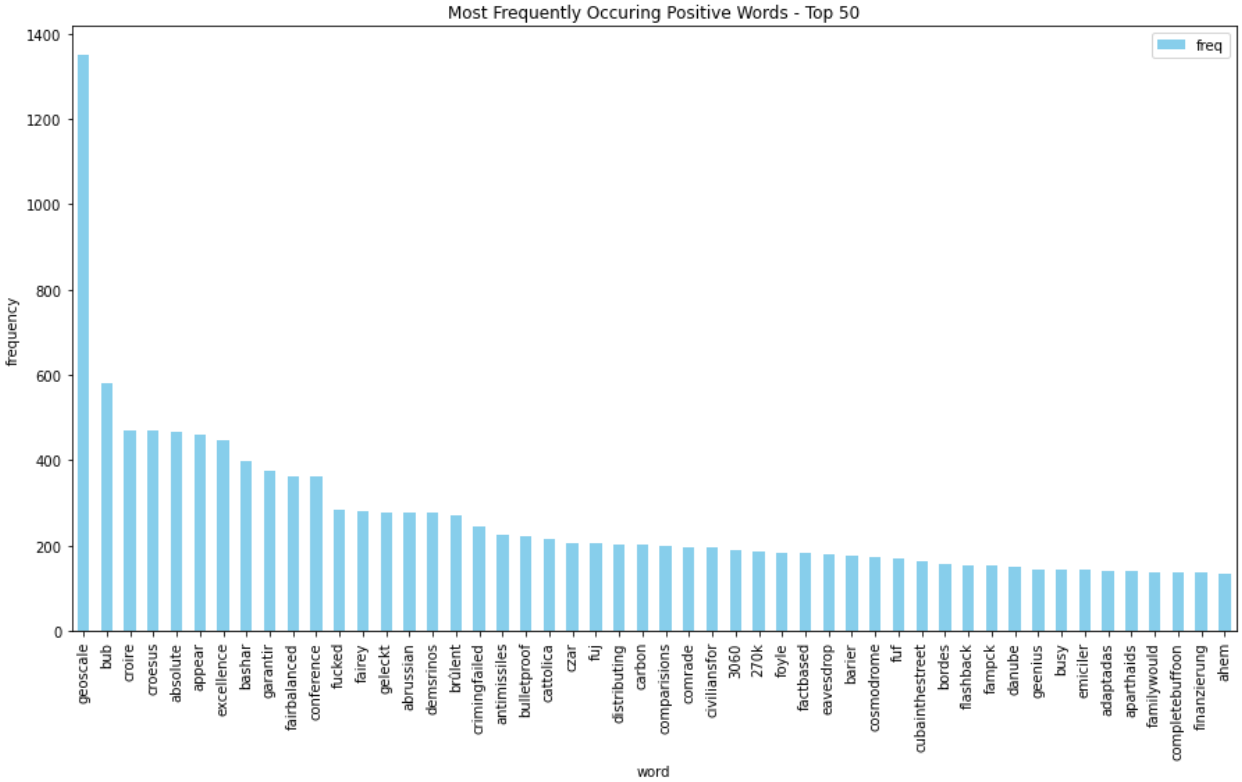

Figure 3. Top 50 Positive Words

In phase 2, we found that most (more than 75%) of the replies to Musk's tweets were negative. In-depth research found that negative tweets are more personal attacks on Musk, which may be related to the following:

1. Entrepreneur meddles in international politics and contradicts his previous support for Ukraine (Starlink and the mentioned tweet).
2. Rumors of directly discussing Ukraine issues with Putin and the Kremlin (cannot be verified).
3. Originally, the public had severely divided views on Musk.

The third suggestion is that Ukraine starts supplying water to Crimea, a concession to Russia. The remaining three proposals agree with Russia's previous conditions for the Russia-Ukraine mediation: Ukraine will not join NATO, denazification, demilitarization and neutralization. In general, it is trying to separate Crimea from Ukrainian territory. Also, "denazification" is likely to be Putin's fabricated rhetoric for Russia's aggression. Musk's statement is irresponsible. Although he has always advertised absolute freedom for the Ukrainian people, Musk's tweet adds rationality to Russia's Invasion. This inconsistency is why netizens showed high aggressiveness towards Musk.

2. Findings in Russia-Ukraine Conflict Twitter Dataset

From phase 3, we got the keywords of negative and positive tweets about this dataset, respectively, which are some words that best reflect netizens' spiritual world. Given that our research aims to advise the Ukrainian government and NGOs to elevate Ukraine's international influence and image, we mainly focus on negative tweets, as negative comments are generally more indicative of problems society faces.
The picture below shows the keywords of the negative tweets we collected. We classified the keywords and found that the tweets reflected four social issues:

- The political and military conflicts between Russia and Ukraine, which we refer to as "CONFLICT".
- We use "ANTI-WAR" to refer to war panic and fierce anti-war sentiment.
- We use "PEOPLE" to refer to people's livelihood issues and the needs of refugees.
- We use "WORLDWIDE" to refer to concerns about political situations around the world.

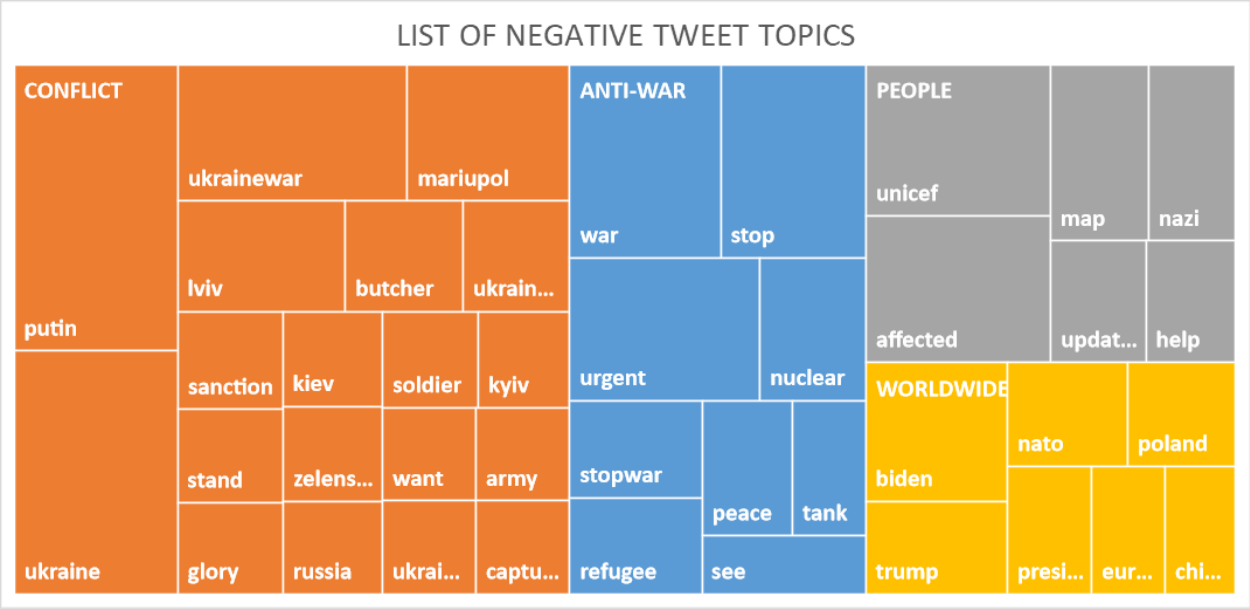

Figure 4. Visualization of the Negative Topics

For the "CONFLICT" part:

- Not surprisingly, "putin" is the word with the highest frequency in this category. As the initiator of the war and the supreme leader of the invader, Putin is widely condemned, and some use "butcher" to describe him. And most think Putin is anti-human and anti-international.
- Many place names appear in this category. In fact, "kiev" and "kyiv" both mean Kyiv, the capital of Ukraine. "mariupol" refers to the city in the southeastern part of Ukraine, Mariupol, which is Ukraine's largest port in the Azov Sea and one of the most important industrial cities in Ukraine. On the first day of the Russian-Ukrainian conflict, on February 24, 2022, Russian troops landed here. "lviv" refers to Lviv, the capital of Lviv Oblast, a major industrial, cultural and educational center in western Ukraine, bordering Poland. Lviv became a vital target of the Russian army after the war began since there used to be a NATO military base.
- The words "stand" and "glory" are derived from two sentences widely quoted by netizens, "Glory to our defenders!" and “Stand with Ukraine”.

For the “ANTI-WAR” part:

- The words in this part can also be divided into two categories. The first is "stopwar", "stop", “peace”, and "war", which are derived from tweets like slogans, aiming to appeal to netizens' anti-war sentiment.
- The second part contains "see", "refugee", and "tank", these types of tweets are mainly from the experience of netizens, describing the destructiveness of war. These tweets promote the anti-war process using bloody reality.

For the “PEOPLE” part:

- "unicef": United Nations International Children's Emergency Fund has been dedicated to reporting on the difficult situation of children and refugees in war. As in all wars, reckless decisions by adults put children in extreme danger. Children cannot be spared in such armed operations. In addition, the destruction of infrastructure, including schools, parks etc., and urban turmoil caused by the war resulted in a crippled childhood. Also, conflicts can have a devastating impact on children: loss of life, disability, detention, recruitment by armed forces or groups, sexual abuse, exploitation or trafficking. Conflict separates families, and thousands of children must provide for themselves and their siblings.
- "updated", and "map" refers to the action of netizens sending real-time updated geographic information on Twitter. For the rest of the world, it might be a better understanding of war, while for the people involved, more information leads to a higher survival chance. It reflects the behavior of the Ukrainian people in helping each other in the face of invaders, and it makes the world aware of the resilience of the Ukrainian people.

For the “WORLDWIDE” part:

- "biden" and "trump" are the two words with the largest proportion in this category. People on Twitter continue to discuss the differences between the two US presidents. Trump himself also posted a lot of criticism speeches. He expressed sympathy for Ukrainians and praised the "courage" of Ukrainian President Volodymyr Zelenskyy. Trump also criticized US President Biden at the rally, saying

Putin took advantage of Biden's "weakness" to launch a military attack when speaking at Conservative Action in Conservative Action Florida on Feb. 26.

- "Poland" indicates the actions of the German Third Reich and the Soviet Union in partitioning Poland in World War II. The presence of the word "Poland" here is evidence of most people's concern about "World War III" and the repetition of history. In another way, this reflects the people's condemnation of the Russian invasion and the desire for peace.
- In this study, we also found an interesting phenomenon. "China" was also mentioned several times on Twitter, but curiously, the content of the tweets themselves had nothing to do with China. After further study, we got some speculative comments and analyses. For example, China persuaded Putin to launch the war after the Winter Olympics which cannot be proved nor denied. However, a takeaway is that the war has such a massive impact on geopolitics that people will believe and spread rumors and cause panic and misunderstanding. This is a sign of panic people losing trust in government authorities. At this time, appeasement from governments and International Organizations is crucial.

3. Recommendations to the Government

In terms of armed conflict: Stop the war requires terms that both Russia and Ukraine agree on. Obviously, this condition cannot be met at this stage. Therefore, the Ukrainian government needs to reduce active conflicts as much as possible and ease the war without apparent concessions. However, this is very hard since the terms from Russia are irrational and violate Ukraine's sovereignty. One feasible way is that Ukraine relies on the support of the United States and European allies and downgrading the war to local conflicts. Depleting Russia's economic and material reserves with minimal losses and thus increasing the bargaining power of peace talks.

On the other hand, after the de-escalation of the conflict, Ukraine should embark on economic recovery and post-war reconstruction to minimize the war's losses for the people. From the analysis, it is also clear that the public has considerable concern and enthusiasm for rebuilding Ukraine, and we believe that the assistance of the international community will not be absent.

In terms of anti-war propaganda: First, the Ukrainian government should actively express its anti-war will and expose Russia's hypocritical nature as an aggressor. We all know that justice must prevail. Such an approach would hinder Russia's military actions, bring Russian soldiers to their senses, and stop them from fetishizing military power. Without the inner support from its army, Russia will soon return to the table with reasonable terms and a mind of peace.

International NGOs should also go to the regions under Russian jurisdiction to propagate the spirit of pacifism. Telling the truth about the suffering the Ukrainian people are experiencing makes them aware of the cruelty and destructive power of war. Also, the families of soldiers should know about the battle's brutality with the aid of war's true nature; they can be mobilized to exert pressure on the government to end this inhumane military aggression.

In terms of addressing "PEOPLE": The Ukrainian government needs to ensure the safety of people's life and properties in the country and restore the normal life of the people. On this basis, it is necessary to ensure the right to food, education, work and essential infrastructures. For example, online courses can solve the problem of insufficient school buildings. The fundamental problem is to prevent large-scale conflicts as well as full-scale wars.

NGOs need to remain active in Ukraine. The main tasks are to provide medical services for innocent casualties caused by fighting, to provide educational resources for children, and to provide transportation and other transportation conditions for refugees. If fighting is inevitable, the Ukrainian government should actively move residents to the rear, far away from the flames of war, and, if necessary, abandon the territory and tighten the defense line to ensure the safety of the people.

In terms of stabilizing the international situation: NGOs and local governments should actively help third-party countries accept refugees, help refugees integrate into the local area, and provide employment opportunities for refugees. In areas with many refugees but lack employment opportunities, NGOs and local governments should help promote infrastructure construction and provide shelter for refugees. At the same time, it

can also create employment opportunities and solicit investment for the city to develop basic industries or agriculture to guarantee the people's basic livelihood. Otherwise, transportation should be arranged to balance the number of refugees according to the carrying capacity of each region.

Also, introducing more international pressure could be a viable means of ending the war. Many European countries are currently suffering from energy shortages due to the energy crisis caused by the Russia-Ukraine war. Skyrocketing energy and food prices are a problem for many countries, and their governments are looking for a solution. Initiating multilateral talks and using the influence of other major powers to put pressure on Russia may force Russia to give in and eventually end the war.

Conclusion

To conclude, according to the sentiment analysis of Twitter data, we see a wide range of opinions on the war between Ukraine and Russia. More than 75% of the replies to Musk's tweets are negative. Despite several reasons accounting for it, the major one is that Musk supports Russia's intention of keeping Crimea. It also reflects that most netizens tend to criticize the Russian invasion of Ukraine and call for peace. Through further analysis of the Russia-Ukraine conflict Twitter dataset, the most frequent keywords of negative sentiment can be classified into four categories: 'conflict', 'anti-war', 'people', and 'worldwide', which covers almost every aspect of netizens' opinions. We can see that the Russia-Ukraine conflict brought worldwide attention, and netizens came up with slogans like "Stand with Ukraine" etc., showing their critical options. At this point, we advise the Ukrainian government to actively seek help from its allies, the United States and Europe, and try to negotiate with the Russian government to find a peaceful way to end the war while advocating anti-war will to optimize its international image. NGOs can address the crisis of people by providing medical services, necessities and available resources for those suffering from the war in Ukraine. Also, NGOs should connect with other countries and facilitate their work of accepting refugees. In the end, wars may seem far away to most of us since the only way we get to know them is through the news. In contrast, it's a matter of life and death for Ukrainians who have lost their home and become refugees. We genuinely hope the war will end soon and strongly condemn any acts of violence.

Reference

Romanko, O. (2022). *Sentiment Analysis* [Unpublished Data Set]. University of Toronto

Romanko, O. (2022). *Opinion of influencers – tweets of Elon Musk and responses to his tweets* [Unpublished Data Set]. University of Toronto

Tariq, M. (2022). *Russia-Ukraine Conflict Twitter Dataset* [Data Set]. Retrieved from: https://www.kaggle.com/datasets/tariqsays/russiaukraine-conflict-twitter-dataset

Appendices

1. Appendix A. Part 1 Visualization

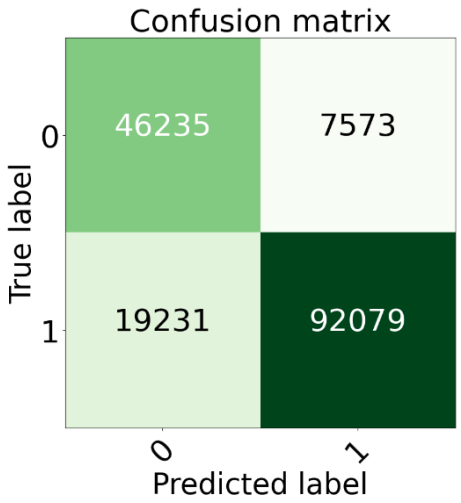

Figure A1. Random Forest

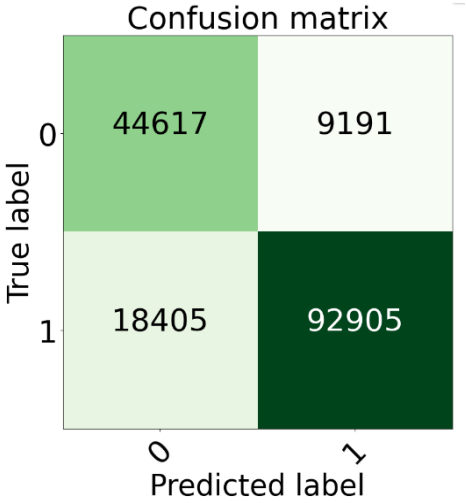

Figure A2. Logistic Regression

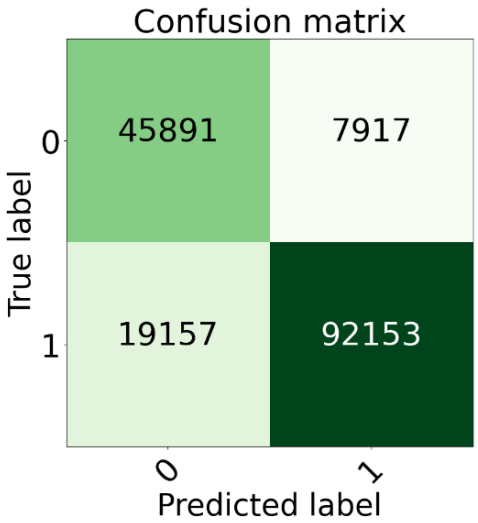

Figure A3. Decision Tree

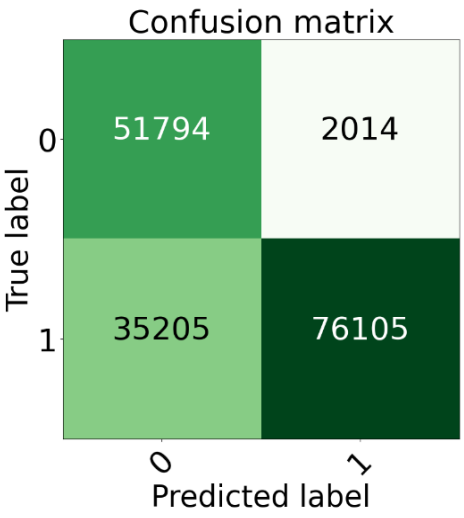

Figure A4. Naive Bayes

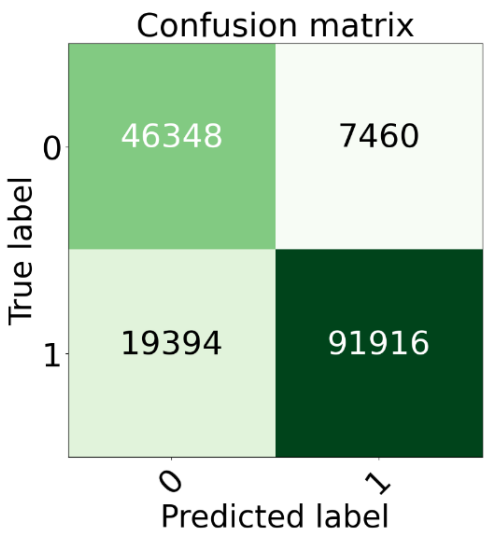

Figure A5. XGBoost

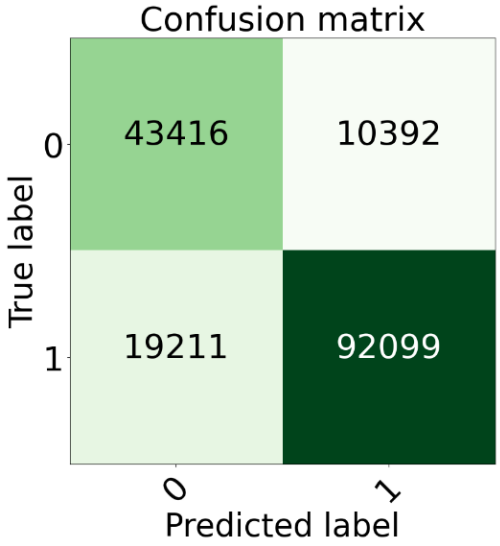

Figure A6. k-NN

2. Appendix B. Part 2 Visualization

Figure B1. Elon's Tweets' Responses

Figure B2. Russia-Ukraine Conflict Twitter Dataset

3. Appendix C. Part 3 Visualization

Figure C1.Word Cloud for All Words

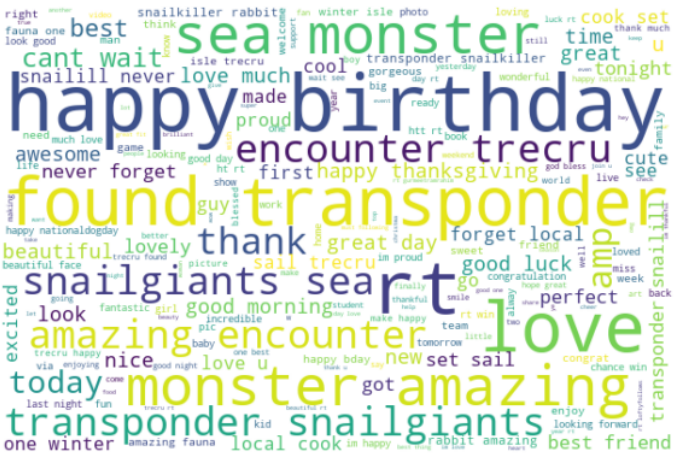

Figure C2. Word Cloud for Positive Words

Figure C3. Word Cloud for Negative Words