1 Executive Summary

Pulse is a social networking service (SNS) data analysis company that focuses on providing social media insights using the latest technologies in artificial intelligence. In this report, we will introduce two proposed products: **Hawkeye** and **ProofReader**, and the insights gathered on the **Russo-Ukraine War** by analysing the most recent SNS information obtained from the microblogging platform Twitter, as well as make tweet recommendations for the Ukrainian government to improve the sentiment for Ukraine in this war.

Hawkeye is built upon an entity-based zero-shot classification model, using the Facebook-built **bart-large-mnli** model on a Happy Transformer framework. Our classification model is able to predict the level and direction of sentiment as well as provide multi-lingual support. Hawkeye's features include multiple visualisations for sentiment trends in topics of interest, word clouds that provide word occurrence in data, and an interactive dashboard tool that allows users to visualise Hawkeye's key topic groups. We have applied this function to the Russo-Ukraine War and will demonstrate how it can be used to understand sudden sentiment changes on a daily basis.

ProofReader is a second product complementing Hawkeye that provides a prediction score with word clouds associated with positive and negative sentiment to aid the user in crafting a tweet or post that will benefit the user without facing the negative implications of having it online. For Ukraine, their government account can leverage this tool to build on supportive sentiments. With the combined tools, the Ukrainian government can output more meaningful tweets, predict public opinion, and recover from unfavourable events that may occur in the future.

With the combined tools, Pulse can help organizations, businesses, and even celebrities have a better real-time understanding of the public opinion and discourse that is occurring and better tailor their communications.

2 Background

Over the last decade, social media has transformed businesses, organizations, self-expression, and even social discourse. Twitter, the world's leading microblogging platform, has opened up communication channels, allowing for real-time public discourse on businesses, products, and global events. Furthermore, Twitter provides a rich source of emotional intelligence, as users frequently express their sentiments and emotions, which can profoundly affect individual behaviours and decision-making.

In 2022, nothing has been more divisive than the Russo-Ukraine conflict. Russia invaded Ukraine in February 2022, resulting in tens of thousands of deaths on both sides, a massive refugee crisis, and food shortages globally. During this time, the world has been divided. Russia has claimed that its goal is "demilitarizing and de-Nazifying Ukraine"[3]. Ukraine has fought back, beseeching foreign government support [7]. Both sides have paid close attention to social media in order to gather and rally support [1]. However, Ukraine has been more innovative and successful at using videos, photographs, and imagery to garner support. Their efforts have led to a strong global response to the war, with protests erupting around the world in support of Ukraine and netizens around the world pressuring local governments to intervene [5]. The ebb and flow of support for both Russia and Ukraine can be encapsulated by public outcries found on Twitter. Every day, over 500 million tweets are sent by more than 300 million active users worldwide [9].

Clearly, having a pulse on public sentiment is a powerful tool for any government, organization, or business to have. The millions of tweets over the course of the war can be studied and analyzed, providing real-time monitoring of the perceptions of each side. Furthermore, this data can be used to predict future sentiments in response to a social media post.

The objective of our business is to help organizations keep a real-time "finger on the pulse" of public sentiments using the latest technologies, including natural language processing (NLP), a discipline of artificial intelligence, and machine learning.

3 How it Works

Our team has developed an algorithm that can comb through Twitter and capture posts about a topic based on specified hashtags that are frequently updated. This dataset of relevant tweets is then preprocessed and analysed using a technique known as sentiment analysis. As proof of the algorithm's efficacy, we have done a case study on the Russo-Ukrainian War. The data is collected by selecting tweets with specified hashtags by the developer (Appendix A). This list of tags is updated frequently to maintain a relevant influx of current events and tweets; more tags are added to the list if they are found on posts where the original hashtag was found.

Data from the month of November was used, up until the 29th of November, and fitted to our model. With this subset of data, our team began the data-cleaning process to prepare for model fitting. As NLP processes all characters, it is important to remove website URLs, emojis, punctuation, and any irrelevant words (stop words) in the context of our data. The removal of stop words is important to eliminate as much "noise" or interference in the data as possible prior to fitting the model in order to increase the prediction accuracy. It is important to consider that the removal of stopwords may contribute to the loss of information, given that some identified stopwords may represent other forms of meaningful data.

Next, the text data is tokenized to break up phrases and sentences from each tweet into individual words, or tokens. Tokenizing the data allows for the construction of a vocabulary within the model that outlines all unique text. Lemmatization of the data follows, which groups together different variations of the same word [4]. This keeps the overall vocabulary of our model smaller and more efficient. Once the data has been stripped down to its relevant data and processed, a spell check function will be used to ensure that all words that the model identifies are correctly spelled, possibly removing more ünique"words from the data.

3.1 What Makes Pulse Different?

Older sentiment analysis services only determine if a tweet is positive or negative; without insight in determining the context of the tweet. Typically, manual translation or further data processing is required to determine the direction of the sentiment presented in each tweet. To put this into context, on average there are approximately 9000 tweets pulled per hour on the Russo-Ukraine War, meaning it is very time-consuming to determine the direction of the sentiment using older sentiment analysis. This also means that the accuracy of specialised topics is low, as these old models are built to be generalized. Given such generalised models, they do not employ multi-language support due to the limitations of their product.

Hawkeye solves all of these issues by using a zero-shot, entity-based, multi-language sentiment analysis protocol. With Hawkeye, government organizations and businesses are able to track the sentiment toward any given topic. For example, in our case study on the Russo-Ukraine war, the visualisations of the trend (**Fig. 1**) allow the user to identify rises and declines in the sentiment trend per day, and specific topics or reasoning can be identified. This feature will allow the user to pinpoint the root cause of the change in sentiment and allow them an opportunity to respond and take appropriate actions. This first product functions as a monitoring system that alerts the user of sudden changes in sentiment. The next question, however, is determining the best possible response to either mitigate sentiment loss or build on sentiment gain. ProofReader is our supplementary product that will generate a report on the favorability score prediction of a given response prior to it being sent out. To aid in generating a better score, ProofReader will also generate word clouds and word importance maps as features to guide the user in creating a better-performing tweet.

4 Algorithm Fitting

An **entity-based zero-shot** classification model is used in our design as the model of choice. The benefit of this classification is the ability to classify data that does not appear during model training as well as identify sentiment based on the provided topics. The model learns to associate the observed classes with the non-observed classes to aid in its ability to provide an accurate prediction. For example, the sentence structure Ï love Ukraineör Ï love Russia"will be identified by older models as both being positive but with no suggestion as to which direction it is polarized. Our model is able to determine that Ï love Ukraine"has positive sentiment directed towards Ukraine and not Russia. Traditional models are only able to identify the magnitude of the sentiment without being able to classify how the text is directed.

Our algorithms use a framework called Happy Transformer, which is built upon the Hugging Face Transformer library. Under this framework, there are many zero-shot classification models to choose from that are tuned for specificity. Thus, this allows us to select from a small sample of models that perform similar tasks and test their performance. Upon testing, our team decided to move forward with a Facebook-built **bart-large-mnli** model, which resulted in the best classification accuracy. Our model requires 6 hours to process a month's worth of data, or approximately 10 minutes to process 1000 samples, which is a very short turnover for analytics and decision-making. This estimate is based on the premium Google Colab runtime; this may vary depending on the computing power of your machine.

5 Visualizations - Dashboard

Visualizations of data is a very useful tool in immediately analyzing trends and public opinion responses to different events. It is also important to understand how these events affect the way the model outputs the data. In **Fig. 1**, three lines are plotted, one for each of the three groups: Ukraine, Russia, and neutral. There are 3 significant days that are displayed in the month of November: the 15th, 18th, and 21st. These sudden changes in data are caused by sudden events that happen. To obtain the topics that were discussed on a specific day, Latent Dirichlet Allocation (LDA) and word clouds were used to find the topics and visualise the frequency.

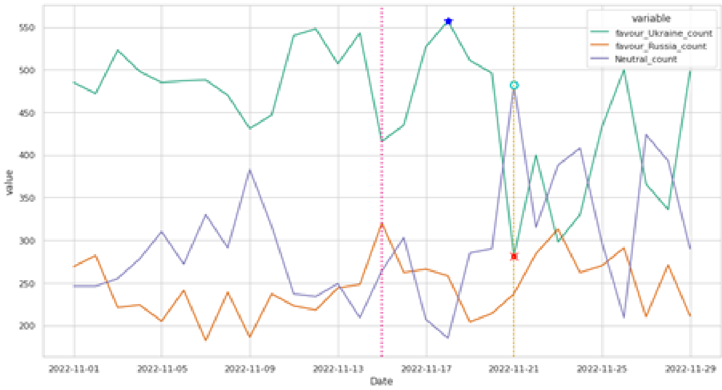

Fig 1: Plot of measured sentiment from tweets for November 2022, notable dates are November 15th, 18th, and 21st.

LDA is a statistical model algorithm that predicts the probability that the text is on a topic based on the text and its relationship to the topic. To put this idea into context, we have three main areas of focus for this product: whether the sentiment is with Ukraine, Russia, or neutral. These tweets have been parsed by the selected Facebook model, but do not provide any information as to what the topic is. LDA is able to sort the information fed into it and provide a probability as to which class it belongs in. Figure 2 depicts how data from each topic are grouped and related using a distance map and term frequency.

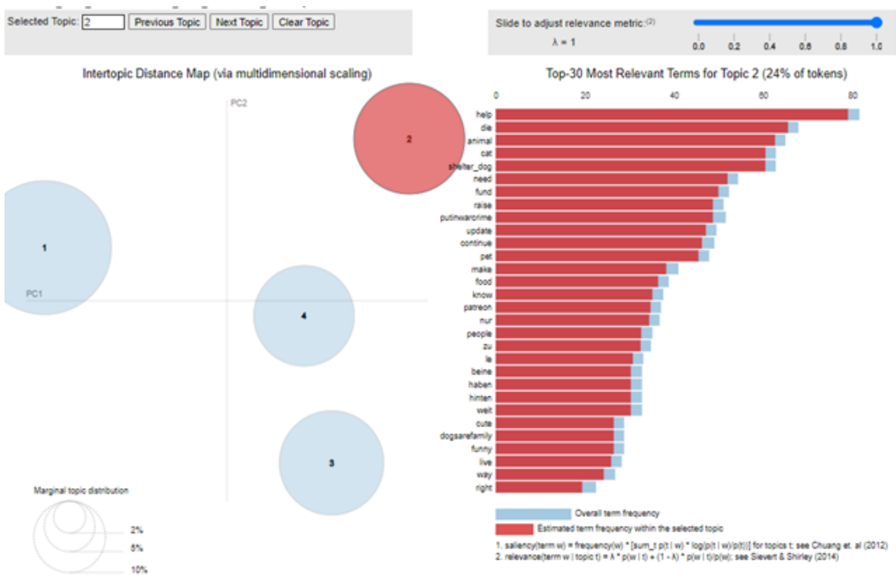

Fig 2: Interactive distance map and dashboard of term frequency (per topic). In order to understand what happened on the 15th, 18th, and 21st, word clouds were made based on the occurrence of unique words that appeared in tweets on those days, along with word clouds associated with 4 topic groups generated by LDA. The results are summarised in **Figs 3, 4, and 5**.

5.1 Visualization - Word Cloud

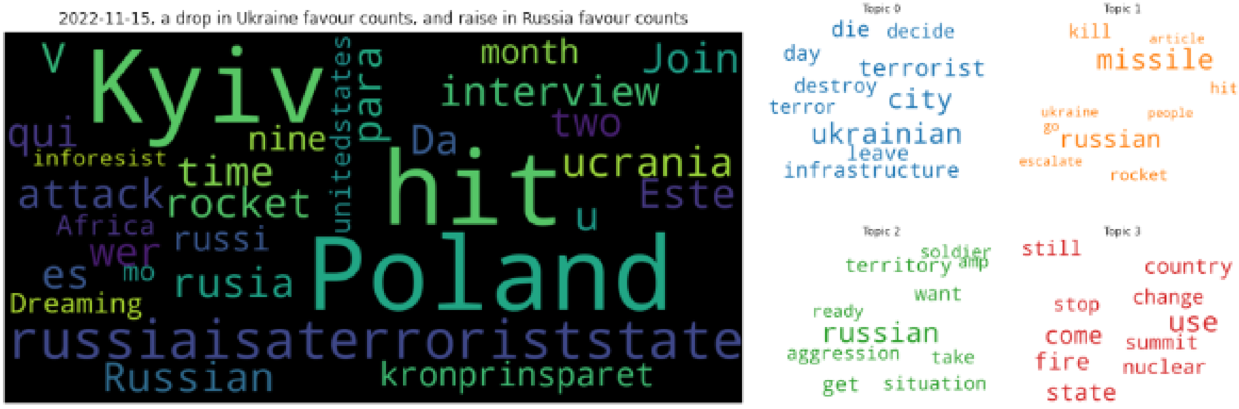

Fig 3: Word clouds of tweets on November 15th.

Fig 4: Word clouds of tweets on November 18th.

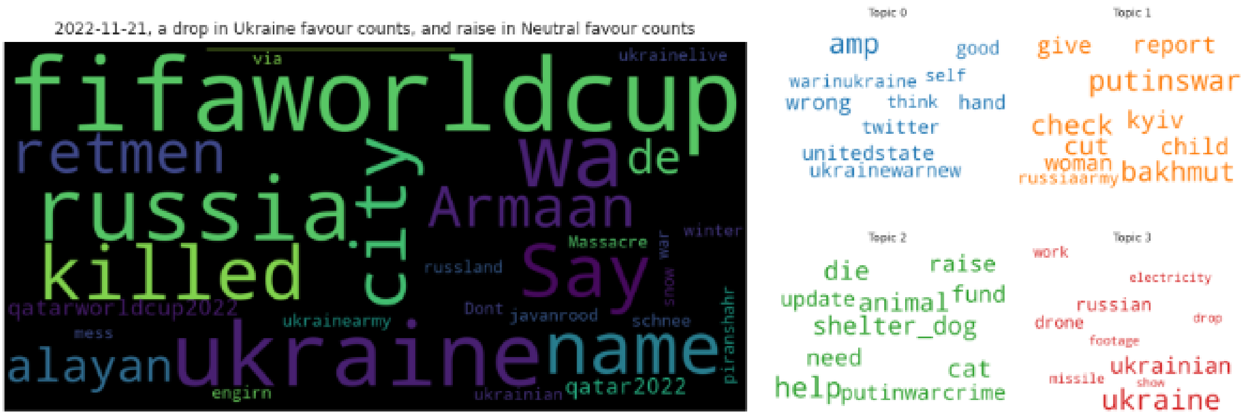

Fig 5: Word clouds of tweets on November 21st.

On November 15th, a stray missile, or rocket, was directed from Ukrainian airspace and made contact with Polish soil. NATO had made a statement that suggested it was of Ukrainian origin, and the idea was quickly shut down by Zelensky, Ukraine's president [6]. As NATO had made a statement regarding this out-of-line behaviour, it makes perfect sense as to why netizens believed that Ukraine had made a terrible decision and directed their discontent towards them online. Ukraine's mood reached an all-time high on November 18th, just three days after the missile strike incident. This may be caused by a multitude of reasons [8]. That was the day that power was restored to many Ukrainians following a Russian strike that had damaged Ukraine's ability to supply and distribute electricity. Additionally, it was the same day that the first train back to Kherson, a city occupied by Russia for 8 months, departed to celebrate the liberation via Ukrainian efforts.

Lastly, on November 21st, the sudden spike of neutrally favourable tweets spiked due to the opening match of the FIFA World Cup. But we can observe that the main topics of discussion still include war-related topics referring to both Ukraine and Russia. The topic word clouds also show that the majority of the topics are grouped to include only relevant topics to the Russo-Ukrainian war.

The appearance of the neutral-favor spike may lead you to believe that the system is flawed or that it is deviating from the target. This, however, demonstrates how the model can classify neutrally favourable data while displaying important war-related topics. The nature of the data collection algorithm is another factor that contributed to the selection of so many keywords on non-war-related words. The data collection algorithm is designed to ”memorize”hashtags that appear alongside the initial specified search targets and hold them for searching for 4 total search iterations [2].

6 ProofReader

Our company offers an add-on product that will proofread the tweets using collected historical data in relation to that topic, as well as aid in word selection by presenting a word cloud of influential words that increase sentiment towards the party of interest. This is done by selecting the main tweet and retrieving the replies under that tweet, then performing entity-based analysis on its replies to generate a score. The replies correspond to the ”audience’s” reaction to the tweet which in turn controls the overall sentiment of a given post.

Our team used 2 sample datasets: one from Elon Musk, an influencer, and one from the official Twitter account for Ukraine, a government representative. This is done to represent our company’s aim to expand this service for celebrities to maintain their image on SNS once the Russo-Ukrainian War concludes. We analysed the tweets for each of the 2 datasets and produced several word clouds that show the words correlated with positive and negative sentiment, which will be part of the final product, as shown in **Figs. 6 and 7**.

These images are generated from Hawkeye data and are further sorted to either positive or negative sentiment under a topic. The second part of ProofReader is to provide a score as to how well the tweet-to-be-sent will perform by analysing the words used. With smaller datasets, as observed for the Elon Musk dataset, it is observed that the predicted score does not perform as expected. However, it is still able to classify if the tweet is going to provide a low or high score. Given more data, the model will perform much better. Larger datasets, provided by the official Ukrainian government, allow for ProofReader to make much more accurate predictions on the performance of a potential tweet; this is shown in the proof of concept with very similar scores obtained by the prediction and the actual score rated by Hawkeye.

Fig. 6: Word clouds associated with single and double word meanings for Elon Musk tweets, positively and negatively associated with sentiment.

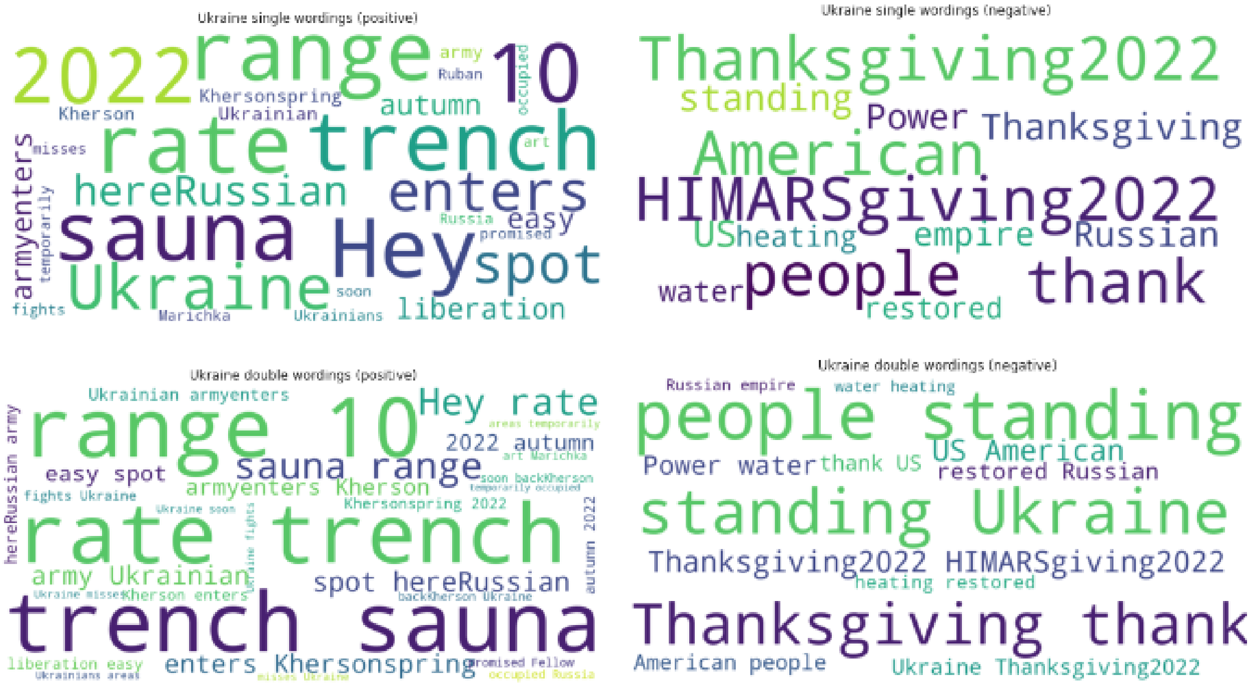

Fig. 7: Word clouds associated with single and double word meanings for Ukrainian government tweets, positively and negatively associated with sentiment.

In **Fig. 8**, we can see a snapshot of the prediction and actual score associated with the positive sentiment of a Ukrainian government tweet. By relating the words used in that tweet with the word clouds in **Fig. 7**, we can see that the Ukrainian government can identify the topics and possible related words that will increase sentiment.

In **Fig. 9**, we can see that Elon Musk's tweet received a very low ProofReader score and received an even lower actual score by Hawkeye. Looking at **Fig. 6**, it is observed that when Elon Musk mentions Crimea or discusses Russia, the sentiment toward his tweet decreases. It is possible that since he mentioned Ükraineïn his tweet, the model scored the tweet higher than it should.

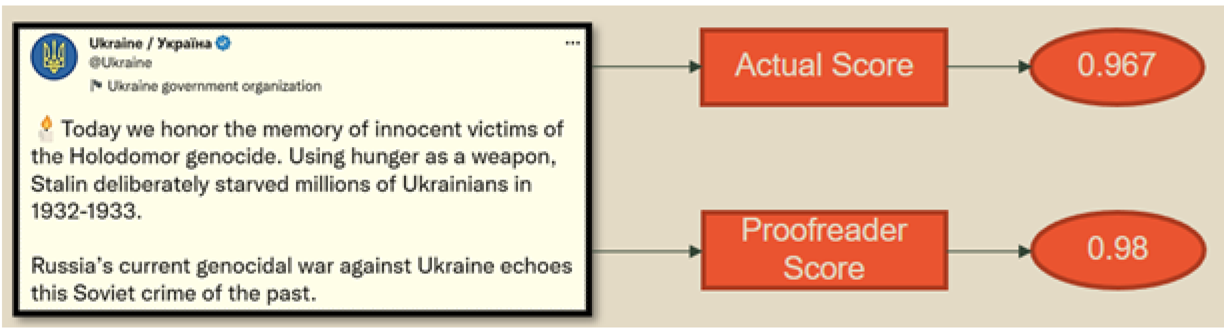

Fig. 8: Visual depiction of the **ProofReader** product extracting a prediction score very similar to the actual score rated by **Hawkeye**.

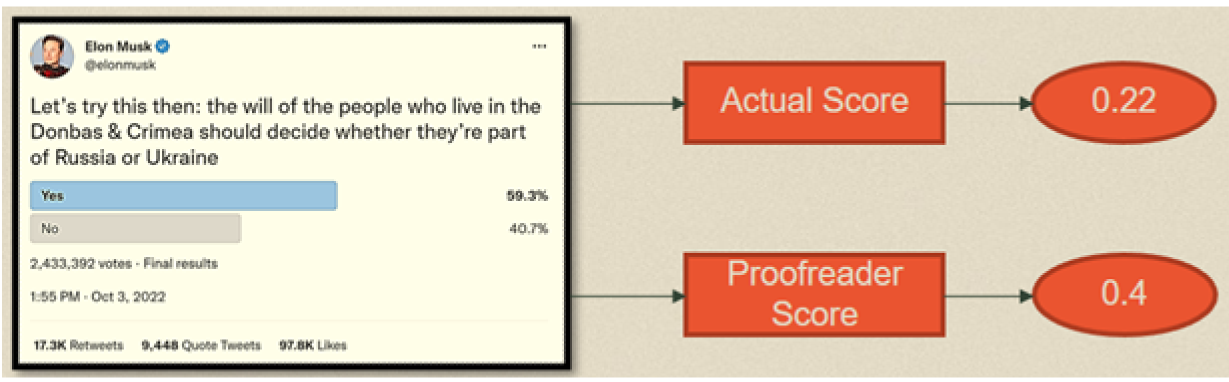

Fig. 9: Caveats of the **ProofReader** when working with smaller datasets, but it is seen that the product is still able to make the classification of a good or bad tweet (threshold = 0.5).

7 Conclusion and Next Steps

It is clear that engagement on social media is paramount for businesses and organizations, and that detecting and affecting sentiment can be a powerful engagement tool with significant consequences. Pulse strives to mitigate possible consequences of negative sentiment trend-setting by pinpointing the cause of the issue through powerful visualization analytics and providing recommendations as to how to address new posts, and ongoing current events. Providing the ability to predict the sentiment and reaction of a possible post gives the client much more information to decide on the viability of their content as well as function as a social safety net. Together, these tools harness artificial intelligence, and the latest technology to give potential clients a greater ability to interact with the public and ultimately strategically communicate their cause.

The next step for Pulse is to examine possible improvements in the model that can aid in better classification and sentiment identification. In our pre-trained model, our team imposed a hypothesis of "This example is label", which is a definite statement. This could be improved by changing the hypothesis to "This example favours label", which shifts the hypothesis statement to only providing the direction of sentiment vs a definite classification of a tweet or post.

8 References

[1] Sara Brown. *In Russia-ukraine war, Social Media Stokes ingenuity, disinformation*. Apr. 2022. URL: https://mitsloan.mit.edu/ideas-made-to-matter/russia-ukraine-war-social-media-stokes-ingenuity-disinformation.

[2] BwandoWando. *Ukraine conflict twitter dataset*. Apr. 2022. URL: https://www.kaggle.com/datasets/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows/discussion/320105?sort=published.

[3] Paul Kirby. *Why did Russia invade Ukraine and has Putin's war failed?* Nov. 2022. URL: https://www.bbc.com/news/world-europe-56720589.

[4] Christopher Manning, Prabhakar Raghavan und Hinrich Schutze. *Stemming and Lemmatization*. 2008. URL: https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html.

[5] *More than 15,000 Russians have been arrested in anti-war protests*. URL: https://www.economist.com/graphic-detail/2022/03/22/more-than-15000-russians-have-been-arrested-in-anti-war-protests.

[6] *Russia-ukraine war: Zelenskiy dismisses NATO's suggestion missile that hit Poland was Ukrainian – as it happened*. Nov. 2022. URL: https://www.theguardian.com/world/live/2022/nov/16/russia-ukraine-war-live-news-emergency-g20-meeting-after-russian-made-missiles-land-in-poland.

[7] Andrea Shalal und David Lawder. *Ukraine's Zelenskiy Appeals for* 55 *bln to cover budget gap and Reconstruction*. Okt. 2022. URL: https://www.reuters.com/world/europe/ukraine-needs-3-4-bln-external-financing-per-month-next-year-imf-chief-2022-10-12/.

[8] Hannah Strange, Sana Noor Haq und Matt Meyer. *November 18, 2022 Russia-ukraine news*. Nov. 2022. URL: https://www.cnn.com/europe/live-news/russia-ukraine-war-news-11-18-22/index.html.

[9] *Twitter usage statistics*. URL: https://www.internetlivestats.com/twitter-statistics/.