TOPICS AND SENTIMENTS ON THE RUSSO-UKRAINIAN WAR

DAAN SCHEERDER

THESIS PROPOSAL SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
OF TILBURG UNIVERSITY

STUDENT NUMBER

2019074

COMMITTEE

dr. Dimitar Shterionov
dr. Koen Haak

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

February 1, 2024

NUMBER OF WORDS

7476

TOPICS AND SENTIMENTS ON THE RUSSO-UKRAINIAN WAR

DAAN SCHEERDER

Abstract

For this thesis, a large dataset of tweets was analysed to find out if a combination of automated topic modelling and sentiment analysis was able to reveal how topics and sentiments on the Russo-Ukrainian war progressed on Twitter during 473 days. Since the start of the full-scale invasion of Ukraine by Russia, policymakers and scientists have been interested in public sentiments on the war. Social media like Twitter can be used as a source of information to get a better view of sentiments in society. NMF proved to be an effective method for topic modelling of tweets. It was capable of finding consistent long-term topics. A combination of three different automated sentiment analysis tools shows varying results. The results from VADER sentiment were mostly negative sentiment, TexBlob resulted in more neutral and positive sentiment, Flair sentiment labels were more negative than both. The methods did however show similarities that proved informative in combination with topic modelling. A combination of Topic modelling and sentiment analysis proved to be effective in modelling discourse on Twitter. Topic modelling allowed tweets to be categorised and analysed as a group, while sentiments indicated how opinionated a topic was. The RBO-score was able to create timelines of similar topics which gave insights into the progression of topics. Combined with the progression of sentiments and topics this gave insights into the discourse on Twitter about the Russo-Ukrainian war.

1 DATA SOURCE, ETHICS, CODE, AND TECHNOLOGY STATEMENT

The data have been acquired from Kaggle BwandoWando (2023) and are freely available to download and use. The creator of the dataset used the official Twitter API to acquire tweets relating to the Russo-Ukrainian war. The micro-blog posts from Twitter are from writers who posted their messages on the Twitter platform. Twitter (now called 'X') informs users how their data can be used, including for research purposes. The analysis does not include the identities of Twitter users.

All figures belong to the author and all code is original work. If code is adapted from other sources it is indicated in the Python notebook available on a GitHub repository and Google Drive. Links to the code and files are on page 15. Python version 3.10.12 was used, together with VADER sentiment version 3.3.2, Flair 0.13.1, TextBlob 0.17.1, nltk 3.8.1, scikit-learn 1.2.2, pandas 1.5.3 and numpy 1.23.5.

This thesis has been written using LaTeX in Overleaf[1] using a template from Tilburg University. To assist with writing, the built-in spellcheck in Overleaf was used together with the free version of Grammarly[2] as a web browser plugin. The latter offers up to a hundred generative AI prompts, that give writing suggestions on sentences. This function was used to improve the author's original work if the suggestions were considered an improvement on the original. More often it was used as an inspiration to reformulate manually. Grammarly was primarily used for spell-checking and grammar. No other generative AI or typesetting tools were used for this thesis.

2 INTRODUCTION

On 24 February 2022, Russia started a full-scale military invasion of Ukraine (Haltiwanger, 2022). The price of energy rose dramatically as a result of the war. The European energy markets were heavily reliant on Russian supplies before the invasion and were heavily affected (Adolfsen, Kuik, Lis, & Schuler, 2022). The war is a humanitarian crisis for Ukraine and caused many Ukrainians to flee to neighbouring countries as refugees (Schwirtz, Kramer, & Gladstone, 2022).

Ukraine is heavily reliant on foreign aid to be able to continue its defensive war against the Russian invasion. This makes it reliant on political and societal support in foreign countries to get the resources it needs. Because public support is crucial in a democratic system, policymakers in Europe are interested in opinions and sentiments about the war to inform their decision-making. For example, the European Commission conducted a survey among a sample of citizens from different EU countries to find out what they think about the war, about the leaders and about the policies, such as sanctions against Russia (Ipsos, 2022). Traditionally, organisations make use of surveys, opinion polls, and focus groups when they need public opinions. With the rise of social media, surveys are becoming less

[1] See: https://www.overleaf.com/
[2] See: https://www.grammarly.com/

relevant. Information on opinions and sentiments is often publicly available on the Web, specifically on social media like Twitter and internet forums. For many organisations, sentiment analysis can replace traditional methods of gathering public opinions (B. Liu, 2012).

Social media can also be used to manipulate or change popular perceptions of the war. Both the Ukrainian and Russian governments try to get support for their causes using social media. The sentiments and subjects discussed on social media can give insights into public support and perception of the war.

With the rise of big data came the demand for analytical techniques like natural language processing (NLP), and topic modelling to find patterns and relations in data, reduce the dimensionality and predict future outcomes (Elragal & Klischewski, 2017).

Social media like Twitter contain huge amounts of postings. Processing, identifying and extracting different opinions and sentiments will be a difficult task for human readers. Therefore, automated sentiment analysis is essential to gain insight into online sentiments (B. Liu, 2012). Similarly, Twitter postings cover different topics and the volume is too large to determine the topics discussed manually. Therefore, automated topic modelling techniques are necessary to gain insight into the topics discussed. Sentiment towards topics related to the Russo-Ukrainian War, change over time, influenced by daily events (Caprolu, Sadighian, & Di Pietro, 2022). While sentiment analysis can provide insights into the sentiments of individual tweets, topic modelling can help to give insight into the overall discourse and topics discussed on Twitter. Merged together they can give a better understanding of how people think about the Russo-Ukrainian war, about the leaders and about the policies and how these change over time.

For this project, a dataset of tweets related to the Russo-Ukrainian War from 27 February 2022 to 14 June 2023 will be used to gain insight into the development of topics and related sentiments over time. This project will refer to Twitter instead of X because the data for this project was collected before the new owner renamed the platform. First, with topic modelling using Non-negative Matrix Factorization (NMF) (D. D. Lee & Seung, 1999), the most prevalent topics over time will be extracted for every day. Second, sentiment analysis using the Valence Aware Dictionary for sEntiment Reasoning (VADER) (Hutto & Gilbert, 2014), TextBlob Loria (2020), and Flair Akbik et al. (2019) will give insight into sentiments of these topics. Rank-biased overlap (RBO) (Webber, Moffat, & Zobel, 2010) will be used as a coherence score between topics. This will be used to track the progression of topics over time. The main topics and related sentiments will be analysed and put into the perspective of important events related to the war.

2.1 *Research question*

> *To what extent does a combination of automatic topic detection and sentiment analysis reveal the progression of topics and related sentiments discussed in tweets about the Russo-Ukrainian war collected from February 2022 to June 2023?*

Sub-questions

1. *To what extent can NMF topic modelling detect the relevant topics in the Twitter discourse about the Russo-Ukrainian war?*
2. *To what extent can VADER, TextBlob and Flair extract the sentiment from Twitter discourse about the Russo-Ukrainian war?*
3. *How can topic modelling and sentiment analysis be jointly used to model the Twitter discourse on the war effectively?*
4. *To what extent do topic modelling and sentiment analysis reveal the progression of topics and sentiments over time?*

3 RELATED WORK

3.1 *Twitter*

Twitter, a micro-blogging platform was, the most used social media platform for analysis (Rathore, Kar, & Ilavarasan, 2017) because it offered easy data collection based on keywords and hashtags using the APIs made available by the platform (Mohamed Ridhwan & Hargreaves, 2021). After new ownership, a lot of the benefits of the platform for research have been phased out. Free access to the API for academic porpoises has been no longer possible since February 2022. This made it impossible to collect data from the platform for this project. Fortunately, since the start of the full-scale invasion of Ukraine many datasets of social media like Twitter have been created, most of them publicly available (Susanne Pohl, Markmann, Assenmacher, & Grimme, 2023). This dataset was created before the rebranding of Twitter into X, therefore this paper will use Twitter to refer to the social media platform.

Two datasets have data from the start of the full-scale invasion of Ukraine in February 2022 until the closing of the Twitter API in June to July 2023. The dataset by Münch and Kessling (2022) contains more than 30 million tweets containing the term #ukraine. The dataset by BwandoWando (2023) also contains more than 30 million tweets but has a broader scope. This dataset also includes tweets that contain other hashtags and keywords. Therefore, the assumption can be made that this dataset is more suitable for this project, where it is important to gain insight into all the main topics discussed on Twitter related to the war. The dataset by BwandoWando (2023) has been used in other literature in the past year. Schwarz, Aranda, and Hartmann (2023) use a part of this dataset for automated situational awareness reporting and Meena and Tokekar (2023) use the data for community detection techniques in combination with sentiment scores. Both use the data to test a certain model or technique, the subject matter of the tweets is of lesser concern. As opposed to previous work with this dataset, this project will primarily focus on the content of the tweets. By combining topic modelling and sentiment analysis, it becomes possible to analyse the discourse and sentiment on Twitter. These can then be placed in the context of the war and important events and give insights into this Twitter dataset that previous work has not revealed.

3.2 *Topics and sentiment on social media*

Since the start of the invasion of Ukraine, Twitter data has been used to gain insight into the topics and sentiments on the war using topic modelling and sentiment analysis techniques. Garcia and Cunanan-Yabut (2022) analyse the sentiments and emotions on Twitter on the first day of the Russian invasion. They concluded that negative sentiments were prevalent and sadness was the most common emotion. Other related work focused on detecting emerging terms and topics on the war, using a combination of co-occurrence information (TF-IDF) and metadata of tweets on Italian tweets on the Russo-Ukrainian war (De Santis, Martino, Ronci, & Rizzi, 2023).

Like this project, previous work has combined sentiment analysis and a form of topic detection to find out how sentiment and topics change over time on social media. For instance, Caprolu et al. (2022) performed an aspect-based sentiment analysis on Twitter related to the Russo-Ukrainian war one month before and after the full-scale invasion and found that for instance sentiment towards Putin became more negative after the invasion. Others analysed a larger period to get some insights into the development and persistence of topics and sentiments over time. Maathuis and Kerkhof (2023) used messages from the social media platform Telegram to analyse the main topics and sentiments in the first two months after the start of the Russian invasion. They use a combination of non-negative matrix factorization with Kullback–Leibler divergence (NMF-KL) (Birjali, Kasri,

& Beni-Hssane, 2021) and Flair (Akbik et al., 2019) sentiment analysis to gain insight into topics and sentiments per day and over an entire period of a 100 days. Similarly, this project will focus on topics and sentiments during the first 473 days of the Russian invasion of Ukraine. The main drawback of Flair is that it does not have a neutral category. This paper will therefore also use other automated sentiment analysis techniques that include a neutral category, which are VADER, TextBlob.

3.3 *Topic modelling*

Other studies on COVID (Jang, Rempel, Roth, Carenini, & Janjua, 2021) or the Russo-Ukrainian war (Caprolu et al., 2022) show that sentiments are topic-sensitive. Therefore, topic modelling techniques are necessary to get a more comprehensive view of the Twitter discourse on the war. Topic modelling is a set of techniques that aim to find hidden topical patterns in a collection of texts. The most popular techniques can be considered Latent Dirichlet allocation (LDA), Latent Semantic Analysis (LSA), and Probabilistic LSA (Egger & Yu, 2022).

Recently, there has been an increase in interest in new algorithms, such as NMF, Top2Vec (Angelov, 2020), and BERTopic (Grootendorst, 2022). In most research, LDA is still used as the standard model. However, especially for short text data from social media, BERTopic and NMF, outperform LDA (Egger & Yu, 2022). This research will use NMF to find the topics in the Twitter discourse. LDA tends to over-generalize topics (Chemudugunta, Smyth, & Steyvers, 2006). Although some find that NMF is less consistent in producing coherent topics than LDA and LSA (Stevens, Kegelmeyer, Andrzejewski, & Buttler, 2012), other research shows that the topics produced by NMF models are clearer and more distinct, compared to other methods like LDA (Egger & Yu, 2022; O'Callaghan, Greene, Conway, Carthy, & Cunningham, 2013). Similarly, Chen, Zhang, Liu, Ye, and Lin (2019) conclude that NMF is more likely to produce topics of higher quality than LDA in short text topic mining tasks using Twitter data. Based on previous research on topic modelling one can assume NMF is the best-suited method for this task.

3.4 *Sentiment analysis*

Sentiment analysis is a Natural Language Processing (NLP) task that aims to extract opinions and sentiments from text data (Birjali et al., 2021; B. Liu, 2015). Previous research often uses VADER to extract sentiment from tweets (Chandrasekaran, Mehta, Valkunde, & Moustakas, 2020; Mohamed Ridhwan & Hargreaves, 2021; Valdez, ten Thij, Bathina, Rutter, & Bollen, 2020). VADER is a lexicon and rule-based sentiment analysis method that is considered a gold standard in finding sentiment in micro-blog text and performs better than other lexicon-based approaches (Bonta, Kumaresh, & Janardhan, 2019). VADER takes into account word order and classifies words based on a word sentiment lexicon that includes negations, contradictions, acronyms, slang and emoticons. These properties make VADER especially suited for sentiment analysis of social media posts. It delivers results on par with more sophisticated machine-learning techniques but has the advantage of being computationally economical (Hutto & Gilbert, 2014). This efficiency means it is especially suited for this project, where millions of tweets need to be classified.

It is not within the scope of this project to manually check whether VADER correctly classifies the tweets. Therefore other sentiment analysis techniques will be used to give more validity to the results.

Besides VADER, other automated sentiment analysis techniques are available. Most of them are not specifically suited for social media messages like those from Twitter and do not take into account emoticons. One method that is specifically trained to detect sentiments in Twitter messages, is the pre-trained sentiment analysis model (Antypas et al., 2023) based on RoBERTa (A Robustly Optimized BERT Pretraining Approach) (Y. Liu et al., 2019) published by Cardiff University's NLP department. However, this model is very resource-intensive, making it slow for large datasets.

Previous work used TextBlob to analyse sentiment on Twitter (Abiola, Abayomi-Alli, Tale, Misra, & Abayomi-Alli, 2023; Chang, Ng, Xu, Guizani, & Hossain, 2022). This lexicon-based sentiment analysis technique seems to have a slight bias towards neutral and positive sentiment compared to VADER(Abiola et al., 2023) as well as human annotation (Aljedaani et al., 2022) and is slightly less accurate than VADER (Bonta et al., 2019). Other than VADER, it is not trained on social media texts and has no emoji recognition which can be a drawback when sentiments are expressed in emoji. Although it has some drawbacks for the analysis of tweets, similar to other lexicon-based approaches, this method is less resource-intensive than more state-of-the-art models based on neural networks. This makes it especially suited to act as an extra sentiment model to add validity to the results of this project.

Birjali et al. (2021) uses the Flair (Akbik et al., 2019) sentiment model to determine the sentiment of Telegram messages about the Russo-Ukrainian war. The Flair sentiment model is trained on an IMDB dataset (Maas et al., 2011) of movie reviews. Although movie reviews are not the same as tweets, they are both relatively short in length. The difference is that reviews are always opinionated and can be considered positive or negative, whereas tweets can also be informative and neutral. Using Flair to classify tweets as either positive or negative, may however give interesting results that differ from the other automated sentiment analysis techniques used in this project. The faster RNN-based version of the Flair sentiment model will be suited to use on this large dataset.

Besides VADER both TextBlob and Flair will be used as sentiment analysis techniques for this project.

3.5 *Literature gap*

Although NMF has shown to be effective for topic modelling of short text, like tweets, the most popular topic modelling methods are still LDA, LSA and Probabilistic LSA (Egger & Yu, 2022). This project will use the more modern NMF method to find topics in the Twitter database. Previous literature on sentiments on Twitter mostly uses one sentiment analysis method (e.g., Caprolu et al., 2022; Garcia & Cunanan-Yabut, 2022; Maathuis & Kerkhof, 2023). This project will be more elaborate and will use multiple sentiment analysis techniques to get a better understanding of sentiments on Twitter. Previous research chose relatively narrow timeframes to analyse specific aspects of the war (e.g., Garcia & Cunanan-Yabut, 2022), this project will however use 473 days of Twitter data. What topics and themes are going to be analysed will be informed by the results of the analysis. This project will also attempt a novel approach to find patterns in topics over time. An RBO score will be used to calculate the similarity between topics on consecutive days. In this way patterns, and similarities between topics can be detected over time.

4 METHOD

The next section describes the methods used for this project. Figure 1 shows the summary of the analytical plan.

4.1 *Data cleaning*

A first exploration of the data showed that it needs to be filtered to remove spam and unrelated subjects. While running the first batch of tweets through the NMF topic modelling, it became apparent that the Tweets contained a few unrelated topics and spam. These unrelated tweets caused the NMF model to make unrelated topics. The topic keywords included antiques, NFTs and Shri-Lanka. To filter out most of the unrelated tweets, the text will be filtered based on keywords. The selection of keywords was done to filter out the tweets that were related to the Russo-Ukrainian war.

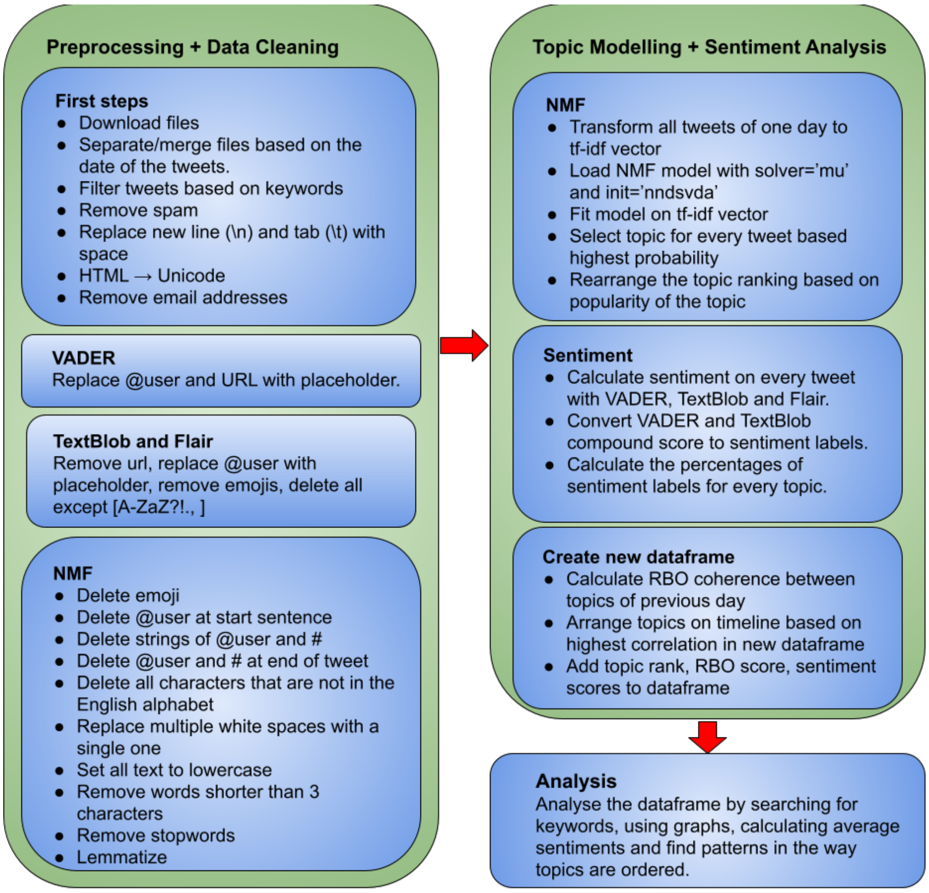


Figure 1: Flowchart of methods.

To filter out tweets related to Ukrainian and Russian leaders and politicians. This includes names of the current and former Ukrainian president and all Ukrainian ministers who were active in the timeframe of the dataset. For Russian politics only the president, prime minister and foreign ministers, to keep the focus on the war in Ukraine. Military leaders of Russia and Ukraine and the former leaders of the Russian mercenary group Wagner are also added to the list which can be found in the code linked in this paper. Next to "Russia" and "Ukraine", it will also use the currencies of both countries and a list of significant names. Furthermore, all names of Ukrainian regions and cities will be included in the list using a database available online (dr5hn, 2023). After a second run of NMF, some spam was still in the topic results which were all messages about the promotion of music. These tweets were easily filtered out with the following keywords: 'amazon', 'now playing' and 'nowplaying'. Because spam was not a big problem after filtering the tweets, no automated spam detection was used.

Next all new line and tab symbols are replaced by a space, email addresses are removed, and HTML code is replaced by Unicode. The preprocessing steps for VADER will include removing user mentions and URL links and replacing them with anonymous placeholders. For Flair and TextBlob the user handles are also replaced by a placeholder, for better comparability with VADER. Next all emojis are removed and words that include numbers. Next all characters that are not upper- or lowercase letters, question marks, exclamation marks, periods or commas are removed. Finally, double spaces are removed.

The text for NMF requires more preprocessing, which will include the removal of excess hashtags and user mentions, stopwords, numbers, punctuation, and unknown characters, as well as lemmatisation and tokenization. In previous work hashtags and user mentions are retained or removed for topic modelling, depending on the subject. Jang et al. (2021) for instance retain user mentions and hashtags because they were deemed informative for the topic modelling and sentiment analysis. User mentions are an @-symbol followed by the user name that directly refers to that user's Twitter account. User mentions are used at the start of a message when it is a response to a specific post. But this is not the only use of user mentions. They are also used in tweets as the subject of a sentence (Kaufmann, 2010). Tweets on the war often include user mentions of important actors in the war, like political leaders and institutions. Indiscriminately removing these user mentions removes important subject information from tweets. For this project, a balance is made between removing excess hashtags and user mentions while retaining the more informative ones that are in the middle of a text. Therefore, user mentions are removed if they are at the start of a tweet or part of a string of multiple mentions, other mentions are retained. Similarly, with hashtags only strings of hashtags are removed, and others are retained. Hashtags or user mentions are at the end of a tweet they are often to label the tweets and not part of the subject of the tweet itself, therefore hashtags and user mentions at the end of a tweet are also removed. For stopword removal, the stop list from the Mallet machine learning for language toolkit (McCallum, 2002) is used which includes 524 common stopwords. After the NMF preprocessing steps all duplicates and tweets shorter than three tokens are removed. After all the previous steps, about 10 million individual tweets remain for analysis. Not all days have equal amounts of data, the first day, 27 February 2022, has about 200k tweets. The number of tweets per day goes down over time, the later days have substantially fewer tweets available for analysis. With most days having more than 10k tweets, but especially the last two months the amount can in some cases be lower than 1000 tweets.

4.2 *Topic modelling*

After the pre-processing steps, the tweets will be converted into a term frequency-inverse document frequency (TF-IDF) matrix, used as input for the NMF model (Egger & Yu, 2022). The TF-IDF matrix will give more weight to more "important" terms.

With NMF, the TF-IDF matrix (D) is decomposed into a product of a document-topic matrix (U) and a topic-term matrix(V). The topic-document matrix (U) consists of probabilities of latent topics for each document. The term-topic matrix(V) consists of term frequencies for each topic (Chen et al., 2019). The Non-negative values of U and V are then iteratively modified to approximate D.

$$D \approx UV \tag{1}$$

Equation 2 shows the loss function. The goodness is evaluated by the square loss between the TF-IDF matrix D and the linearly combined U and V matrices.

$$min \sum_{n=1}^{N} ||d_n - \sum_{k=1}^{K} u_k v_{kn}||_2^2 \quad \text{or} \quad min||D - UV||_F^2 \tag{2}$$

The multiplicative update rule 3 (D. Lee & Seung, 2000) will be used as the learning algorithm, where the two equations below will converge to the final term-topic and topic-document matrices.

$$U \leftarrow U \frac{DV^T}{UVV^T} \qquad V \leftarrow V \frac{U^T D}{U^T UV} \tag{3}$$

Because topic modelling is a form of unsupervised learning, the ideal amount of topics (k) is unknown beforehand. To determine the best value of k, first many NMF models are built based on a range of k topics. The ideal amount of topics for an NMF model can be determined Topic Coherence-Word2Vec (TC-W2V) measure (Islam, 2019; Stevens et al., 2012). A varying number of topics per day would further complicate the tracking of topics over time and add more processing time because the NMF model would need to run multiple times for each day. Therefore this project will get 10 topics per day and use them to create 10 timelines on which the topics can be tracked over time and linked together using an RBO score. This ensures that every topic can be linked to a topic on the prior day based on the highest RBO scores.

The NLTK package in Python will be used for preprocessing and sklearn for TF-IDF and NMF.

4.3 *Sentiment analysis*

VADER will calculate a compound score and classify each tweet as positive, negative, or neutral. Because it uses human-validated rules and a sentiment lexicon, it does not need training data. VADER will calculate compound scores between -1 and +1 based on the sum of the valance score of every word in the document and relevant rules. A score between -0.05 and 0.05 is considered a neutral sentiment score, above 0.05 is positive and below -0.05 is negative (Hutto & Gilbert, 2014).

Similar to VADER, Textblob [3] is a lexicon-based approach for sentiment analysis. TextBlob is a Python library that provides a simple API for many natural language processing tasks, including sentiment analysis, part-of-speech tagging and translation. Other than VADER, TextBlob is not specifically designed for social media texts and does not have emoji recognition. All words in a text get a semantic score based on the sentiment score of the words in a pre-trained dictionary. The semantic score of a text is then defined by the mean score of individual sentiment scores assigned to the individual words. This dictionary is trained using the Sentiment Polarity Dataset from the NLTK library Loper and Bird (2002). The resulting TextBlob sentiment score will be between -1 and 1, with 0 a neutral sentiment, more than 0 a positive sentiment, and less than 0 a negative sentiment.

Instead of the regular Flair (Akbik et al., 2019) sentiment model based on a deep neural network used in Birjali et al. (2021), this project will use the faster Flair Recurrent Neural Network (RNN) version of Flair to save resources and time. This version has an accuracy of 96.83 on the Flair

[3] https://textblob.readthedocs.io/

testing data, just 2.04 points less accurate than the regular model. The Flair sentiment model works differently than the lexicon-based approaches, which calculate a score based on a sentiment lexicon and rules. The RNN sentiment model predicts the sentiment of a text as either positive or negative with a certain level of confidence and is trained on an IMDB movie review dataset (Maas et al., 2011). RNNs are a type of neural network containing a hidden layer that enables it to remember information about sequences. This makes them especially suited for natural language tasks like sentiment analysis.

The topic sentiment for both the lexicon and RNN-based approaches can be calculated by using percentages op of positive, negative or neutral labels. These percentages can then be used to calculate average sentiments for topics that persist over multiple days. A benefit of this approach is that percentages of sentiment labels make lexicon-based approaches comparable to the sentiment labels that result from the Flair RNN. It also provides more information about sentiment distribution in a topic compared to a single average compound score.

4.4 *Topics over time*

To get an idea of how topics on Twitter evolve, the topics that come forward on one day had to be compared with the day before. To compare the lists of topic words one has to take into account the ranking of the words. A form of weighting needs to be applied so the first words in the topic should have a higher weight than lower-ranking ones. Another aspect of comparing ranked lists is that most topics are nonconjoint, meaning not all words of one topic are also in the other. Webber et al. (2010) introduce the rank-biased overlap(4) (RBO) that is especially suited for this purpose. It is top-weighted and takes into account the nonconjointness of the ranked lists. RBO is a weighted overlap measure, that takes a weighted average of the overlap of two lists at increasing depths. The weight is adjusted with the parameter p, the smaller the value of p, the steeper the decline in weights over the increasing depth (d), and the more top-weighted it becomes. To select the right value of p one can look at the progression of weights over every increase in depth and to what extent the first topic words are more important than the last ones. When looking at the average weights for the words for every topic calculated by a first run of NMF, the first words are about 4 times more important than the second words and about 15.6 times more important than the tenth words in the topics. The best value for p will be estimated on the basis that the first word will be 15.6 times more important than the tenth, leading to p=0.737. In figure 2 the average weight distribution of words in NMF topics are plotted against

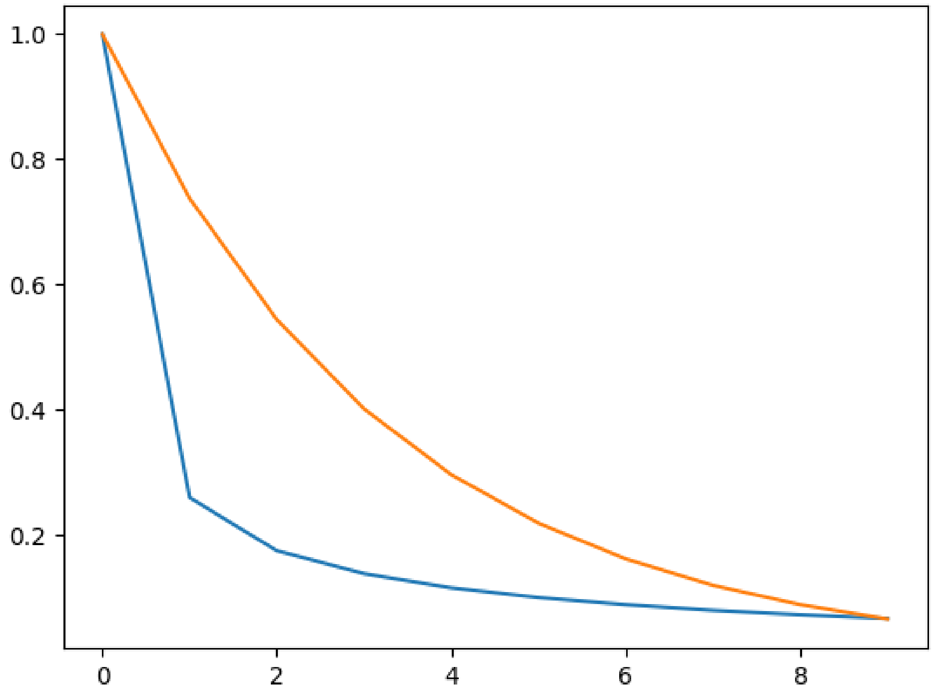


Figure 2: Average weight distribution of words in NMF topics (blue) vs. weight decline in RBO measure (orange).

the weight decline in RBO measure with a p=0.737. Although the weights of words in NMF show a steeper decline, the RBO does approximate the decline in weights in a similar way.

$$RBO(S, T, p) = (1 - p) \sum_{d=1}^{\infty} p^{d-1} \cdot A_d \tag{4}$$

The RBO measure will be used to find out which topics remain constant over several days. For every day in the Twitter dataset the NMF algorithm will create 10 topics, these topics can then be correlated with the topics from the day before. The topics will be linked to each other based on the height of the RBO scores. This will create 10 timelines that will show the progression of topics, their mean sentiment score and how the topics correlate to the topics of the previous day. These timelines can then be used for further analysis and experimentation.

4.5 *Prediction*

To make sense of the resulting timelines it is possible to use manual labeling and analysis to select the important topics and trends. This is a time-consuming task, so it is interesting to find automated solutions for it. As this paper is primarily focused on the analysis of tweets and not on comparing various models and methods, it is interesting to let an automated process find a way to automate this process. van Kuppevelt et al. (2020) introduce Mcfly, a wrapper around the Keras API (Chollet et al., 2015) in Tensorflow (Abadi et al., 2015). Mcfly uses random-search on a range of plausible settings to propose a model architecture and hyperparameters that perform best on the data van Kuppevelt et al. (2020). The data resulting from the RBO coherence between the topics, the topic ranking, and the average sentiment scores can be used to train a model to classify trending, regular and non-important topics. Similar to the activity classification of accelerometer data, these time series together could point to a certain class of topic. To find the right model using Mcfly and train it there needs to be data to train on. This means that a section of the timelines will be labelled for training purposes.

The code and files used in this project will be available on Google Colab or GitHub [4]. On Google Colab more graphs and results are visible in the notebook and can be recreated using the code. The files on Google Drive are available too [5].

5 RESULTS

Topic modelling using NMF extracted 10 main topics for every day in the data set. Topics are represented by a ranked list of 10 topic words. After the topic extraction, the RBO is calculated between days to track the progression of similar topics.

On average VADER and Flair sentiment analysis show a majority of the topics and individual tweets are negative, while TextBlob results in more positive and neutral topics and tweets (table 3). Table 1 shows the accuracy of other sentiment analysis techniques compared to VADER. The sentiment labels resulting from VADER and Flair are the same for 53% of the tweets, while for VADER and TextBlob this is 49%. Flair and TextBlob only share 40% of their labels. When looking deeper at the difference between VADER and TextBlob, the difference is primarily on the negative labels, only 37% of the negative labels in VADER are also negative in TextBlob, whereas Flair and VADER share 69% of the negative labels.

A quick look through the data frame shows that topics with the same first word persist over time. Using the RBO score these topics were linked to each other based on the highest scores. Usually, when topics on two

[4] See: https://colab.research.google.com/drive/1rMocUz5jcAxvzgFaStVnz5bryvfvGBou?usp=sharing or https://github.com/DaanScheerder/Thesis-second-submission/

[5] See: https://drive.google.com/drive/folders/1sAcgCsVKsmtjQvgUzXYPPk8hyoygkooU?usp=drive$_link$

Table 1: VADER sentiment labels compared to those by TextBlob and Flair. The percentages are rounded.

	TextBlob		
VADER	Positive	Negative	Neutral
Positive	62	14	24
Negative	32	37	30
Neutral	29	17	54
	Flair		
	Positive	Negative	Neutral
Positive	58	42	0
Negative	31	69	0
Neutral	40	60	0

Table 2: Example of topics and RBO scores with topic starting with Russian

2022-05-03	russian,army,soldier,military,propaganda,killed,troop,destroyed,tank,invasion
RBO-score	0.162
2022-05-04	propaganda,beware,russian,smell,wake,psyop,lie,unnderstan,played,make
2022-05-03	mariupol,azovstal,civilian,plant,steel,evacuated,evacuation,city,azov,woman
RBO-score	0.478
2022-05-04	russian,mariupol,azovstal,plant,civilian,force,steel,city,region,missile

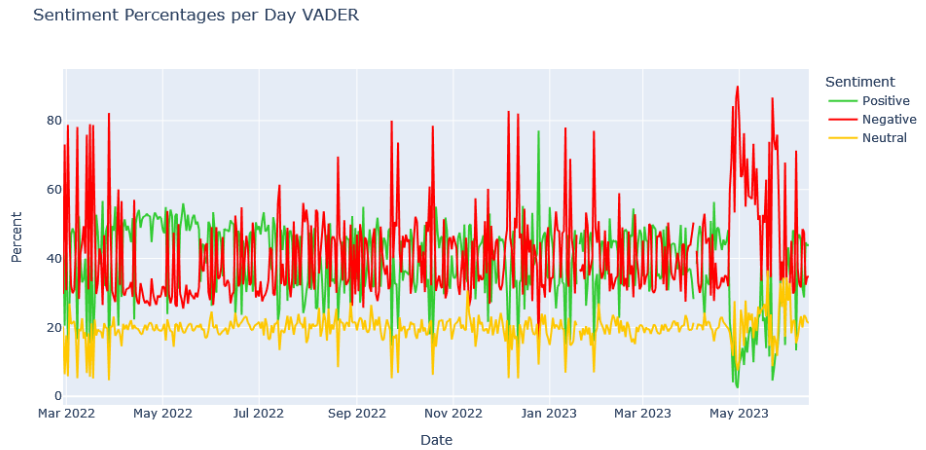
consecutive time points share the same first word, they are linked because of the weight of the first word. However, when the rest of the topic words do not match this may not be the case. One example is seen in table 2 where the topic with Russia as the first word is more correlated to the topic of Mariupol.

A search through the dataset reveals that there are six of these long-term topics. The most popular topic has "Ukraine" as the first word. This topic is a constant on Twitter for 471 days out of 473 days in the dataset and is the most popular topic for 304 days. The ten words that describe this topic are Ukraine, support, invasion, weapon, day, NATO, country, news, military, and aid. These topic words indicate that the topic is primarily about support for Ukraine with military aid, like weapons, or NATO support. On average, this topic is negative with VADER and Flair sentiment, Texblob shows more positive and neutral sentiment 3. Compared to the average sentiment of topics it is more positive with all sentiment analysis techniques, and it is also the most positive out of all the long-term topics.

In 417 out of 473 days, the word "support" or "stand" is present in the topic. Both words refer to a topic that is primarily about supporting

Table 3: Average sentiment distribution of all tweets and average sentiment of all topics. Followed by 6 long-term topics on Twitter. Sentiment labels are in rounded percentages.

	VADER		
Topics	%positive	%negative	%neutral
Average Sentiment Tweets	33	48	18
Average Sentiment Topics	31	51	19
Ukraine, support, invasion, weapon	39	41	20
Russian, soldier, army, troop	34	47	20
War, crime, end, criminal	27	56	17
Putin, Vladimir, president, Trump	29	52	18
Russia, China, sanction, country	32	49	19
Ukrainian, soldier, army, force	28	54	18
	TextBlob		
	%positive	%negative	%neutral
Average Sentiment Tweets	42	26	33
Average Sentiment Topics	40	26	34
Ukraine, support, invasion, weapon	44	24	33
Russian, soldier, army, troop	41	25	33
War, crime, end, criminal	39	27	34
Putin, Vladimir, president, Trump	40	27	33
Russia, China, sanction, country	41	26	33
Ukrainian, soldier, army, force	39	27	34
	Flair		
	%positive	%negative	
Average Sentiment Tweets	42	58	
Average Sentiment Topics	42	58	
Ukraine, support, invasion, weapon	45	55	
Russian, soldier, army, troop	41	59	
War, crime, end, criminal	41	59	
Putin, Vladimir, president, Trump	40	60	
Russia, China, sanction, country	42	58	
Ukrainian, soldier, army, force	42	58	



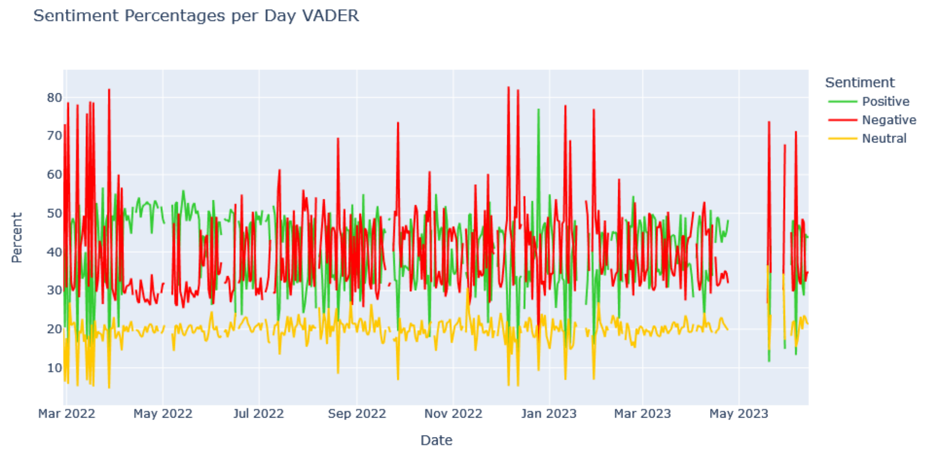


Figure 3: Difference in VADER sentiment graphs before (top) and after (bottom) filtering out topics that do not contain "stand" or "support"

Ukraine or in other words "stand with Ukraine". For 336 days "support" is the second word of the topic. Especially towards the end of the Ukraine topic timeline, between 24 April and May 19 2023, the topic changes to news and updates about Ukraine, including news about Soledar, a settlement on the front line during that time. During this period an increase in negative sentiment was visible with all sentiment analysis methods. When filtering out all days that do not contain these two words, the average sentiment changes to 40% positive, 40% negative and 20% neutral, making it even more positive. Figure 3 shows the difference in VADER sentiment on this topic before and after filtering on support and stand.

One interesting date in the support Ukraine topic is December 25, 2022, which stands out as extremely positive for this topic, with 77% positive with VADER, 81% with Flair and 62% with TextBlob. One explanation for this outlier in positivity is that on this day the president of Ukraine addressed the world with a Christmas speech and it is the first time Ukraine officially celebrated Christmas on this date. This could have triggered people to voice their support for Ukraine. This is signified by Zelensky showing up as the 5th topic word on this date.

Another popular topic on Twitter is about the Russian army. The topic with "Russian" as the first word is present in 469 out of 473 days, and it ranks 67 days as the most popular topic and 257 days as the second most popular. The first ten words that describe this topic are Russian, soldier, army, troop, military, tank, destroyed, force, missile, and region. These indicate that this topic is about the Russian military. When looking deeper into the topic words, the individual topics all include one or more words referring to the Russian military. Including force, military, soldier, tank, troops, army, attack or missile. Compared to the topic about supporting Ukraine, this one is more negative in all three sentiment analysis techniques (figure 3.

The topic of calling an end to the war and Russian war crimes is a third recurrent one present for 462 out of 473 days. The words describing this topic are war, crime, end, criminal, stop, "russiawar", footage, day, win, and peace. The sentiment is the negative on average than the other topics. During a majority of 270 days, more than 50% of tweets on this subject are labelled negative using VADER. An example of how the sentiments on this topic change over time using the three methods is added in Appendix A (page 28). The trajectories of sentiment show some similarities, especially when looking at the outliers. On October 29, there is for instance very positive with all techniques,

Another recurring topic starts with Putin, the president of Russia. It is present for 461 days in the dataset and the ten words describing it are Putin, Vladimir, president, Trump, stop, time, Biden, fuck, west, and leader. This topic is slightly less negative than the previous one in VADER and TextBlob, but the most negative in Flair. It primarily is a topic about Vladimir Putin, the president of Russia, but also includes Trump and Biden, the former and current president of the United States. The swearword shows this is a very polarised topic, in line with the more negative sentiment.

The topic starting with Russia and followed by China seems to be about these countries in combination with sanctions. These topic words are followed by country, state, NATO, invasion, west, terrorist, and Iran. The last long-term topic begins with Ukrainian and is mainly about the

Ukrainian military. Described by the words: Ukrainian, soldier, army, force, troop, military, child, refugee, Nazi, position.

Another way to find topics is to use a combination of the RBO scores and the first words of the topics. If the first word is present for two or more consecutive days and linked on the timeline, it is considered a single recurring topic. When filtering out topics that remained for more than two weeks, four new topics emerged from the dataset. Two of those are about the front-line cities, Mariupol and Bakhmut. A third is about Zelensky, the president of Ukraine, and a fourth is about tanks. The topic with Bakhmut as the first word is frequently recurring in the NMF topics and three times it is present for longer than two weeks in a row. The first time for 19 days between 26th of February and the 16th of March 2023, with 54% negative in VADER. It rises from the 9th to the second most popular topic on the4th March after which it goes back to the 7th position and disappears from the timeline. The second time the topic on Bakhmut appears for 27 days between the 21st of March and the 16th of April 2023, and a third time for 49 days between the 18th of April and the 5th of June 2023. These last two instances of the Bakhmut topic are negative for 59% of the Tweets using VADER. The last instance of the Bakhmut topic is the most popular of the three, with 27 out of 49 days in the top 3 topics. The topic words are quite similar, all starting with Bakhmut and Wagner, the Russian mercenary group that besieged the city.

One can also look for topics that are not necessarily the most recurring but are of interest for other reasons, for instance, more time-specific news items, like the Bucha massacre. One way to find the progression of topics like this is to search for topics that have Bucha as the first word, but to broaden the search the first 3 words are chosen. Topics involving Bucha came up on 2 April and disappeared from the topics after 9 April 2022. The topic is most popular at April 4th when it is the 3rd most popular, and it progresses to be the least popular on the 9th, after which Bucha disappears from the topics. Figure 4 shows a positive spike in the sentiment for April 8th, a closer look shows that the topic for that day starts with child and civilian and has Bucha as a third topic word. The sentiment combined with the topic indicates that this is probably a different topic.

6 DISCUSSION

The analysis explored multiple methods to analyse the vast amount of topics and sentiment. It showed what is possible with data from topic modelling and sentiment analysis. The methods proved to be complementary, and together they gave insight into millions of tweets on the Russo-Ukrainian war.

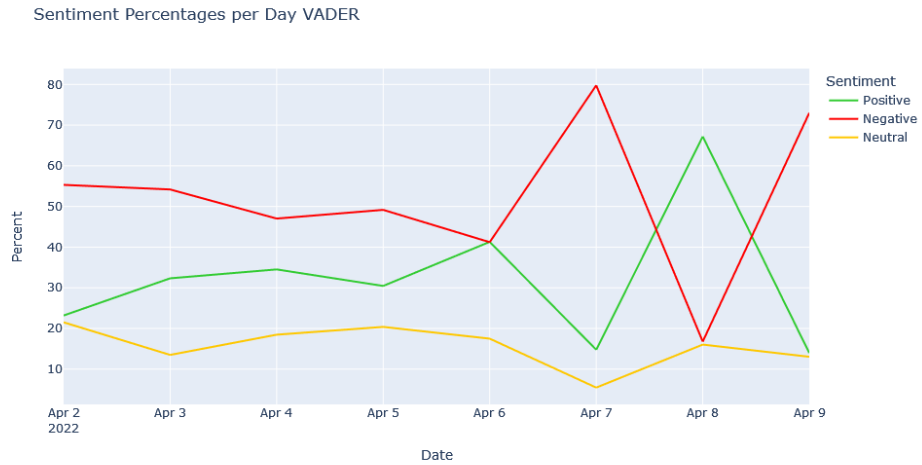


Figure 4: VADER sentiment on Bucha topic

With NMF it was possible to automatically extract recurring topics and emerging topics about newsworthy events in the war. NMF was very effective in picking up relevant news items, like the Bucha massacre. It appeared on April 2nd 2022, right at the moment the first images from Bucha were shared through social media before it became global news on April 3rd. NMF was also able to extract topics that remained constant for almost all days in the dataset. Even though the long-term topics showed constancy in topic words, there can be substantial nuances between these topics. For instance, the topic starting with Ukraine mostly included topic words referring to support, but there were instances where it was more about general news updates about the front line. Further filtering on keywords showed relevant changes in sentiments.

Combining the RBO-linked topics with a search for topic words that remained the same over a period of time on one of the timelines, proved to be successful in extracting emerging and relevant topics. For example, the appearance of the topic on Bakhmut was in line with news on Bakhmut during that time, where Russian Wagner forces captured the city centre in April 2023 Psaropoulos (2023). The RBO score offered a way to organise large quantities of topics to make it more manageable to get a general idea about patterns in the data. These patterns then could inform further decisions on how to proceed in the analysis of topics.

Both VADER and Flair showed that topics were more negative on average. Similar to other studies using TextBlob, this study showed that TextBlob has a bias towards a neutral or positive score. Aljedaani et al. (2022) shows it is more neutral and positive than human annotated sentiment, and Abiola et al. (2023) similarly show that TextBlob results in more neutral and positive scores than VADER. Although Flair missed the neutral sentiment category, it showed more similarity with the results of VADER than TextBlob did, although Flair did show a bias towards negative sentiment compared to VADER.

Averaging out the sentiment for topics over several days results in sentiments that are very close to the overall average of the sentiment in the entire dataset. Because negative emotions tend to spread faster and wider than positive emotions on Twitter (Sliwa., Kušen., & Strembeck., 2023), it was to be expected that sentiments tended to be more negative on Twitter. While there were no truly positive topics, some topics were more positive when compared to the average sentiment. Comparing to the average may thus be a more informative way to check whether a topic is more negative or positive, considering that sentiment is always on the negative side on Twitter.

One drawback of this project is that the amount of topics is fixed at 10. This was done to make tracking topics over time more straightforward but may have resulted in less meaningful topics. According to Islam (2019), the most meaningful and interpretable topics are found by choosing the maximum coherence score at k topics using a coherence measure like Topic Coherence-Word2Vec (TC-W2V). Future research could thus expand on this work, by finding the most ideal amount of topics per day.

The prediction task was not completed for this project, the labelling of data proved to be too labour-intensive to do within the time frame. Future research can focus on further automatisation of the analysis of topics and sentiments on social media over time. Mcfly remains an interesting tool to automate the creation of models that could be trained to find patterns in data and could be used on the data produced in this project.

7 CONCLUSION

This project attempted to analyse a large dataset of tweets to find out if a combination of automated topic modelling and sentiment analysis was able to reveal how topics and sentiments on Twitter progressed during the 473 days of the Russo-Ukrainian war.

NMF topic modelling was effective in extracting topics from the Twitter dataset. It showed that part of the Twitter discourse on the war revolved around six topics, that remained present in the data for almost all days in the dataset. Although topics were not always comparable over long periods due to changes in topics, NMF proved to be very consistent in getting topics out of the Twitter data. Using RBO-scores periods were revealed where topics remained constant over several days, like the topic on Bakhmut. It was also possible to search for well-known topics on the

Russo-Ukrainian war, like the Bucha massacre. This shows that NMF can find topics on Twitter that are relevant and in line with discourse in traditional media, like newspapers.

The three different sentiment analysis techniques showed varying results. Because a human evaluation of the sentiment results was not part of this analysis, there is no way of knowing which sentiment analysis technique was the most accurate. Based on previous research it can be expected that the results from VADER are closest to reality. TextBlob is known to be on the Neutral side and Flair missed the important neutral category. TextBlob and Flair however did serve as an extra validity check for the results of VADER. VADER and Flair were similar for more than half the tweets, despite the lack of a neutral category. Although average sentiments were not similar for topics, the sentiment analysis techniques showed similar trajectories over time. More extreme outliers of positive and negative sentiments for a topic are often visible with all three methods.

The combination of sentiments and topics proved to be effective in modelling discourse on Twitter. Topic modelling allowed tweets to be categorised and analysed as a group, while sentiments indicated how opinionated a topic was. Topic words were also effective in explaining outliers in sentiment over time. The topic words often indicated what was the cause of a sudden change in sentiment on a topic. Instead of a change in sentiment for a similar topic, changes in sentiments can thus be a sign that topics change. Future research can further explore how topic words change over time and how that relates to changes in sentiments.

Topics and sentiments could be combined to reveal two aspects of the discourse on Twitter. The analysis went into different approaches to analyse the topics and sentiments. The long-term topics showed interesting variations and developments in sentiments and topic ranking showed how popularity increases and decreases over time. The RBO-score was able to create timelines of similar topics which gave insights into the progression of topics, which could be combined with the progression of sentiments and topics to get an idea of the discourse on Twitter about the Russo-Ukrainian War.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Retrieved from https://www.tensorflow.org/ (Software available from tensorflow.org)

Abiola, O., Abayomi-Alli, A., Tale, O. A., Misra, S., & Abayomi-Alli, O. (2023, January). Sentiment analysis of COVID-19 tweets from selected hashtags in nigeria using VADER and text blob analyser. *Journal of Electrical Systems and Information Technology*, *10*(1), 5.

Adolfsen, J. F., Kuik, F., Lis, E. M., & Schuler, T. (2022). The impact of the war in ukraine on euro area energy markets. *ECB Economic Bulletin, Issue 4 / 2022*, 46-53. Retrieved from https://www.ecb.europa.eu/pub/economic-bulletin/html/all_releases.en.html

Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019). FLAIR: An easy-to-use framework for state-of-the-art NLP. In *NAACL 2019, 2019 annual conference of the north american chapter of the association for computational linguistics (demonstrations)* (pp. 54–59).

Aljedaani, W., Rustam, F., Mkaouer, M. W., Ghallab, A., Rupapara, V., Washington, P. B., ... Ashraf, I. (2022). Sentiment analysis on twitter data integrating textblob and deep learning models: The case of us airline industry. *Knowledge-Based Systems*, *255*, 109780. Retrieved from https://www.sciencedirect.com/science/article/pii/S0950705122009017 doi: https://doi.org/10.1016/j.knosys.2022.109780

Angelov, D. (2020). *Top2vec: Distributed representations of topics.* arXiv. Retrieved from https://arxiv.org/abs/2008.09470 doi: 10.48550/ARXIV.2008.09470

Antypas, D., Ushio, A., Barbieri, F., Neves, L., Rezaee, K., Espinosa-Anke, L., ... Camacho-Collados, J. (2023). Supertweeteval: A challenging, unified and heterogeneous benchmark for social media nlp research. In *Findings of the association for computational linguistics: Emnlp 2023.*

Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, *226*, 107134. Retrieved from https://www.sciencedirect.com/science/article/pii/S095070512100397X doi: https://doi.org/10.1016/j.knosys.2021.107134

Bonta, V., Kumaresh, N., & Janardhan, N. (2019, Mar.). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, *8*(S2), 1–6. Retrieved from https://ojs.trp.org.in/index.php/ajcst/article/view/2037 doi: 10.51983/ajcst-2019.8.S2.2037

BwandoWando. (2023). *(sunset) ukraine conflict twitter dataset.* Kaggle. Retrieved from https://www.kaggle.com/dsv/5934908 doi: 10.34740/KAGGLE/DSV/5934908

Caprolu, M., Sadighian, A., & Di Pietro, R. (2022). *Characterizing the 2022 russo-ukrainian conflict through the lenses of aspect-based sentiment analysis: Dataset, methodology, and preliminary findings.* arXiv. Retrieved from https://doi.org/10.48550/arxiv.2208.04903

Chandrasekaran, R., Mehta, V., Valkunde, T., & Moustakas, E. (2020, Oct 23). Topics, trends, and sentiments of tweets about the covid-19 pandemic: Temporal infoveillance study. *J Med Internet Res*, *22*(10), e22624. Retrieved from https://doi.org/10.2196/22624 doi: 10.2196/22624

Chang, V., Ng, C. Y., Xu, Q. A., Guizani, M., & Hossain, M. A. (2022). How do people view COVID-19 vaccines: Analyses on tweets about COVID-19 vaccines using natural language processing and sentiment analysis. *Journal of Global Information Management (JGIM)*, *30*(10), 1–29.

Chemudugunta, C., Smyth, P., & Steyvers, M. (2006). Modeling general and specific aspects of documents with a probabilistic topic model. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems* (Vol. 19). MIT Press. Retrieved from https://proceedings.neurips.cc/paper/2006/file/ec47a5de1ebd60f559fee4afd739d59b-Paper.pdf

Chen, Y., Zhang, H., Liu, R., Ye, Z., & Lin, J. (2019). Experimental explorations on short text topic mining between lda and nmf based schemes. *Knowledge-Based Systems*, *163*, 1-13. Retrieved from https://doi.org/10.1016/j.knosys.2018.08.011

Chollet, F., et al. (2015). *Keras.* https://keras.io.

De Santis, E., Martino, A., Ronci, F., & Rizzi, A. (2023). An unsupervised graph-based approach for detecting relevant topics: A case study on the italian twitter cohort during the russiandash;ukraine conflict. *Information*, *14*(6). Retrieved from https://www.mdpi.com/2078-2489/14/6/330 doi: 10.3390/info14060330

dr5hn. (2023). *Countries states cities database.* Retrieved 2023-11-24, from https://github.com/dr5hn/countries-states-cities-database

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in Sociology*, *7*.

Elragal, A., & Klischewski, R. (2017). Theory-driven or process-driven prediction? epistemological challenges of big data analytics. *Journal of Big Data*, *4*(1), 1–20.

Garcia, M. B., & Cunanan-Yabut, A. (2022). Public sentiment and emotion analyses of twitter data on the 2022 russian invasion of ukraine. In *2022 9th international conference on information technology, computer, and electrical engineering (icitacee)* (p. 242-247). doi: 10.1109/ICITACEE55701.2022.9924136

Grootendorst, M. (2022). *Bertopic: Neural topic modeling with a class-based tf-idf procedure.* arXiv. Retrieved from https://arxiv.org/abs/2203.05794 doi: 10.48550/ARXIV.2203.05794

Haltiwanger, J. (2022, January 24). *Russian president vladimir putin announces military assault against ukraine in surprise speech.* Retrieved October, 10, 2022, from https://www.businessinsider.com/putin-announces-military-assault-against-ukraine-in-surprise-speech-2022-2

Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216-225. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Ipsos. (2022). *Eu's response to the war in ukraine : report*. European Union. Retrieved from https://europa.eu/eurobarometer/surveys/detail/2772

Islam, T. (2019). Yoga-veganism: Correlation mining of twitter health data. Retrieved from https://arxiv.org/abs/1906.07668 doi: 10.48550/ARXIV.1906.07668

Jang, H., Rempel, E., Roth, D., Carenini, G., & Janjua, N. Z. (2021, Feb 10). Tracking covid-19 discourse on twitter in north america: Infodemiology study using topic modeling and aspect-based sentiment analysis. *J Med Internet Res*, *23*(2), e25431. Retrieved from https://doi.org/10.2196/25431 doi: 10.2196/25431

Kaufmann, M. (2010). Syntactic normalization of twitter messages.. Retrieved from https://api.semanticscholar.org/CorpusID:7915427

Lee, D., & Seung, H. S. (2000). Algorithms for non-negative matrix factorization. In T. Leen, T. Dietterich, & V. Tresp (Eds.), *Advances in neural information processing systems* (Vol. 13). MIT Press. Retrieved from https://proceedings.neurips.cc/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788-791.

Liu, B. (2012). Sentiment analysis: A fascinating problem. In *Sentiment analysis and opinion mining* (pp. 1–8). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-031-02145-9_1 doi: 10.1007/978-3-031-02145-9_1

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press. doi: 10.1017/CBO9781139084789

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., … Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*. Retrieved from https://arxiv.org/abs/1907.11692

Loper, E., & Bird, S. (2002). *Nltk: The natural language toolkit.* arXiv. Retrieved from https://arxiv.org/abs/cs/0205028 doi: 10.48550/

ARXIV.CS/0205028
Loria, S. (2020, Apr). *Simplified text processing¶.* Retrieved from https://textblob.readthedocs.io/en/dev/#
Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011, June). Learning word vectors for sentiment analysis. In D. Lin, Y. Matsumoto, & R. Mihalcea (Eds.), *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 142–150). Portland, Oregon, USA: Association for Computational Linguistics. Retrieved from https://aclanthology.org/P11-1015
Maathuis, C., & Kerkhof, I. (2023). The first two months in the war in ukraine through topic modeling and sentiment analysis. *Regional Science Policy & Practice*, *15*(1), 56-74. Retrieved from https://rsaiconnect.onlinelibrary.wiley.com/doi/abs/10.1111/rsp3.12632 doi: https://doi.org/10.1111/rsp3.12632
McCallum, A. K. (2002). *Mallet: A machine learning for language toolkit.* (http://mallet.cs.umass.edu)
Meena, S. S., & Tokekar, V. (2023). Sentiment-based community detection using graph transformation. In S. Das, S. Saha, C. A. Coello Coello, & J. C. Bansal (Eds.), *Advances in data-driven computing and intelligent systems* (pp. 65–76). Singapore: Springer Nature Singapore.
Mohamed Ridhwan, K., & Hargreaves, C. A. (2021). Leveraging twitter data to understand public sentiment for the covid-19 outbreak in singapore. *International Journal of Information Management Data Insights*, *1*(2), 100021. Retrieved from https://doi.org/10.1016/j.jjimei.2021.100021
Münch, F. V., & Kessling, P. (2022, Jul). *Ukraine twitter data.* OSF. Retrieved from osf.io/rtqxn doi: 10.17605/OSF.IO/RTQXN
O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2013). An analysis of interactions within and between extreme right communities in social media. In M. Atzmueller, A. Chin, D. Helic, & A. Hotho (Eds.), *Ubiquitous social media analysis* (pp. 88–107). Berlin, Heidelberg: Springer Berlin Heidelberg.
Psaropoulos, J. (2023, April 6). *'they were all shot': Russia accused of war crimes as bucha reveals horror of invasion.* Retrieved January, 15, 2024, from https://www.aljazeera.com/news/2023/4/6/wagner-advances-in-bakhmut-as-ukraine-gears-up-for-counterattack
Rathore, A., Kar, A., & Ilavarasan, V. (2017, 11). Social media analytics: Literature review and directions for future research. *Decision Analysis*, *14*. doi: 10.1287/deca.2017.0355
Schwarz, K., Aranda, D. A., & Hartmann, M. (2023). Towards automated situational awareness reporting for disaster managementmdash;a case

study. *Sustainability*, *15*(10). Retrieved from https://www.mdpi.com/2071-1050/15/10/7968 doi: 10.3390/su15107968

Schwirtz, M., Kramer, A. E., & Gladstone, R. (2022, March 7). *Humanitarian crisis worsens for ukrainians trapped in russia's onslaught.* Retrieved 12 October 2022, from https://www.nytimes.com/2022/03/07/world/europe/ukraine-humanitarian-crisis-russia.html

Sliwa., K., Kušen., E., & Strembeck., M. (2023). An analysis of twitter communities related to the 2022 war in ukraine. In *Proceedings of the 8th international conference on complexity, future information systems and risk - complexis* (p. 27-36). SciTePress. doi: 10.5220/0011980700003485

Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). Exploring topic coherence over many models and many topics. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning* (p. 952–961). USA: Association for Computational Linguistics.

Susanne Pohl, J., Markmann, S., Assenmacher, D., & Grimme, C. (2023, Jun.). Invasion@ukraine: Providing and describing a twitter streaming dataset that captures the outbreak of war between russia and ukraine in 2022. *Proceedings of the International AAAI Conference on Web and Social Media*, *17*(1), 1093-1101. Retrieved from https://ojs.aaai.org/index.php/ICWSM/article/view/22217 doi: 10.1609/icwsm.v17i1.22217

Valdez, D., ten Thij, M., Bathina, K., Rutter, L. A., & Bollen, J. (2020). Social media insights into us mental health during the covid-19 pandemic: Longitudinal analysis of twitter data. *J Med Internet Res*, *22*(12), e21418. Retrieved from https://doi.org/10.2196/21418 doi: 10.2196/21418

van Kuppevelt, D., Meijer, C., Huber, F., van der Ploeg, A., Georgievska, S., & van Hees, V. T. (2020, July). Mcfly: Automated deep learning on time series. *SoftwareX*, *12*.

Webber, W., Moffat, A., & Zobel, J. (2010, nov). A similarity measure for indefinite rankings. , *28*(4). Retrieved from https://doi-org.tilburguniversity.idm.oclc.org/10.1145/1852102.1852106 doi: 10.1145/1852102.1852106

APPENDIX A

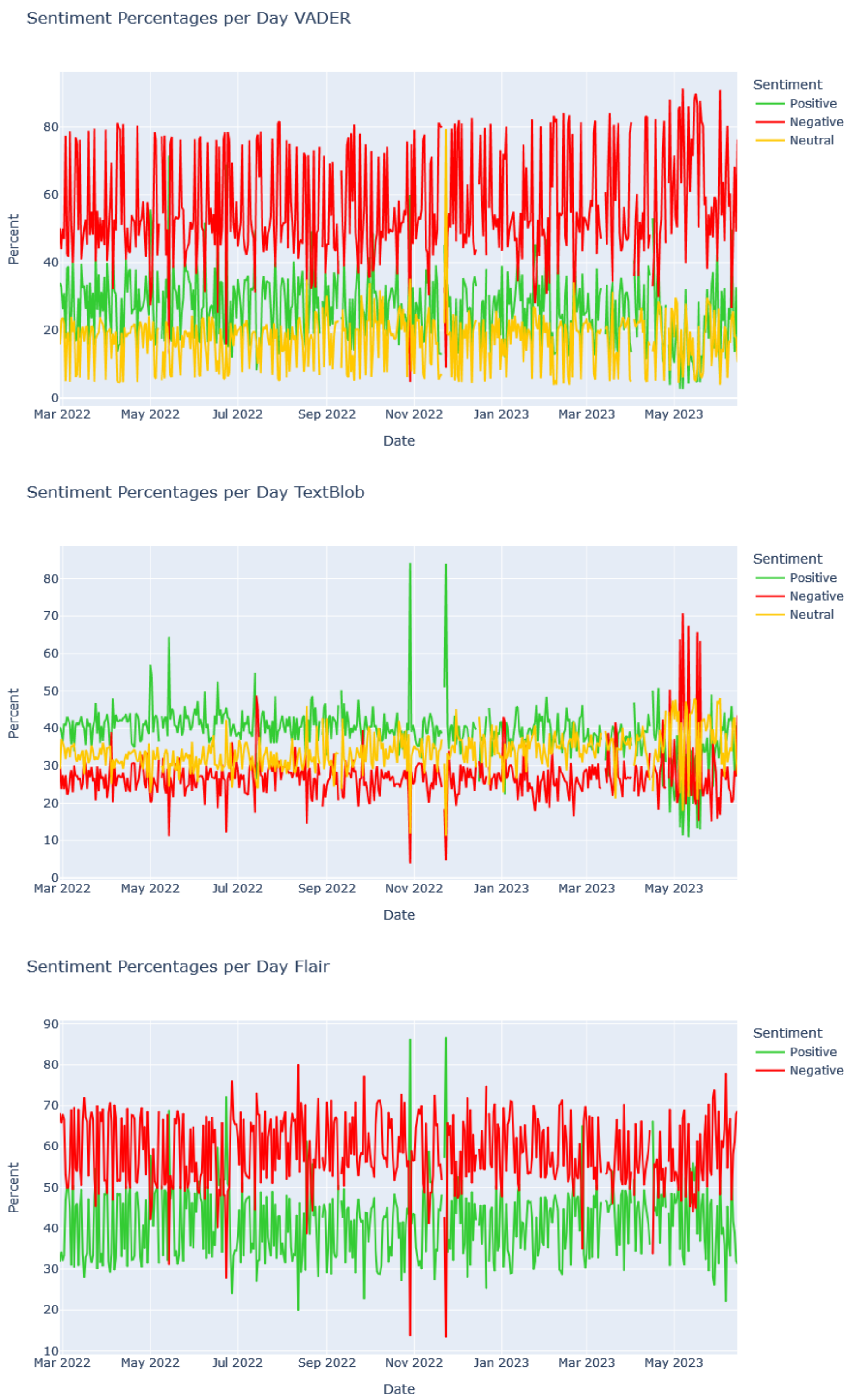


Figure 5: Sentiments on topic about war and war crimes with VADER, TextBlob and Flair