Analysis of Pro-Russian Tweets during Russian Invasion of Ukraine

Josip Katalinić
Department for information and communication sciences,
Faculty of Humanities and Social Sciences, University of
Zagreb, Zagreb, Croatia
josipkatalinicandroid@gmail.com

Abstract **– With the Russian invasion of Ukraine on February 24, 2022 a new aspect of the conflict opens up, the fight on social networks. By analyzing the most influential pro-Russian Twitter profiles that cover the daily events and impacts of the Russian invasion of Ukraine as well as outside it, the number of followers, people following as well as tweets, time series comparison of publishing, sentiment through polarity and subjectivity, and the classification of individual tweets as malicious by building an SVM (Support Vector Machine) classifier were collected and analyzed. For the purpose of training the SVM classifier, a sample from dataset of 3 million identified malicious tweets deleted by Twitter after being linked to the Russian IRA agency known for operating malicious user accounts, available through the Kaggle data science community, was used. The results of the work show that pro-Russian Twitter profiles have clear and defined influence operations that oscillate through different time periods as a reaction to the dynamics and development of the conflict, and thus certain events become narrative elements that can influence the emotions of the target audience.**

Keywords - twitter; text classification; invasion of Ukraine; data analysis; influence operations;

I. INTRODUCTION

With the Russian invasion of Ukraine on February 24, 2022, a new aspect of the conflict opens up, the fight on social networks. An increased number of pro-Russian Tweets showing significant influence operations aimed at directing readers thinking and behavior toward narratives that suit the Kremlin government. In the first two weeks of the conflict, there was a 381% increase in tweets with the context of Putin rather than Zelensky [1].

Reference [2] states that psychological warfare is part of the information war that the Russian Federation uses with the help of "soft power". Reference [3] defines "soft power" as the power of attraction, where an agenda that is considered legitimate is set. While in contrast "hard power" is the use of force, payment and setting a certain agenda based on them. Hard power is push, while soft power is pull i.e., hard power is like waving a carrot or stick, while soft power is like a magnet. In this paper based on collected pro-Russian tweets we are presenting in which way tweets are disseminated on Twitter platform.

In this paper, pro-Russian Tweeter profiles, which contribute to the pro-Russian agenda with their reactions and announcements, were collected and processed. The aim of this research is to observe and present the trends that characterize the collected tweets, and to process the mentioned tweets at the level of sentiment analysis through subjectivity and objectivity, as well as through positive and negative. In addition to sentiment analysis, tweets were also classified as malicious or not based on a previously defined set of malicious users. All the mentioned features are presented through the passage of time, and connected with the significant events of the Russian invasion of Ukraine extending the understanding of influence operations and its patterns on Twitter.

The remainder of the paper is structured as follows. Process of data collection is presented in Section 2. Time analysis of collected tweets is covered in Section 3. Sentiment analysis as well as time series visualization is outlined in Section 4. SVM malicious user classifier results are obtained in Section 5. Study findings are presented and discussed together with future research in Section 6.

A. *Advent of milbloggers*

In addition to the aforementioned influence operations, daily reporting on events in Ukraine as well as events related to the conflict using OSINT (Open-source intelligence) sources enables public policy think tanks [4], as well as private users [5] to create their reports. Access to OSINT makes it possible to define the term milblogger, which refers to a creator who addresses the topics and dynamics of the conflict.

While directing the narrative towards Russia's successes in Ukraine, pro-Russian milbloggers also criticize certain decisions of the Russian military command, such as in the case of Alexander Zhychkovskiy and Alexander Khodarkovsky. During which Zhychkovskiy criticized the neglect of reservists on the front in the Zaporozhye region, which lost priority. Zhychkovskiy reported that Russian commanders trapped their own lightly equipped infantry units in areas of intense Ukrainian artillery fire without significant artillery support, and that they did not rotate other units through the areas to replace them. While Khodarkovsky alleges that Russian commanders are not sending reinforcements in time, preventing Russian forces from resting between attacks [6].

One of the most influential pro-Russian millblogger is Rybar, who has over 1.1 million subscribers on the Telegram channel [7], while his English Twitter profile has over 45,000 followers at the time of writing this paper [8]. Reference [9] states that the founder of Rybar, which employs at least 10 people, is 31-year-old military translator Mikhali Zvinchuk. He is a former employee of the press service of the Ministry of Defense of Russia, born in Vladivostok, studied at the Military University in Moscow, specializing in the Arabic language. From 2015-17, he was employed at the Ministry of Defense and helped organize press trips to Syria for Russian journalists. In addition to transmitting daily news on the field and very detailed war maps, Rybar also serves as a channel for sensitive messages from Moscow. So, Rybar was the first to publish the video of the Russian "nuclear train" that caused panic headlines in Europe [10].

Milbloggers also provide information on which to gather pro-Russian narratives that are part of influence operations. Their reporting, as well as the coverage of events, can therefore create narrative elements that aim to confuse the reader combining high interactivity and the proliferation of the Twitter platform by reaching numerous users.

II. DATA COLLECTION

Using snowball sampling by going to Rybar Twitter followers, 19 active pro-Russian Twitter profiles were selected, which, including the Rybar profile, makes a set of 20 profiles.

TABLE I. OVERVIEW OF SELECTED 20 PRO-RUSSIAN TWITTER PROFILES

Following	Followers	Tweets
82,197	855,124	323,424

For the selected 20 Twitter profiles, tweets were collected from May 6, 2022 until January 22, 2023. The Python programming language and the Tweepy library, which provides access to the Twitter v2 API (application programming interface), were used to collect tweets [11]. While for access to historical data, Academic Research access was used, which was obtained from Twitter for the purposes of writing the paper. From the collected tweets, exclude filters were defined on retweets and replies so that only the author's tweets were collected. This approach aims to prevent a large volume of tweets that are not related to the author.

A. *Analysis of collected tweets*

In addition to collecting the tweet text, the following were also collected: tweet id, publication date, quote count, retweet count, reply count and like count. The number of collected tweets for the specified period is 13,381 tweets.

TABLE II. OVERVIEW OF THE COLLECTED 13,381 TWEETS

Quotes	Retweets	Replies	Likes	Impressions
67,515	718,468	279,683	3,765,282	126,233,781

By analyzing the collected tweets, we can see that 13,381 tweet posts are responsible for 126,233,781 impressions, which indicates a significant influence and reach of each individual posts. While other parameters such as quotes, retweets, replies and likes indicate a high rate of interaction with posts. Each individual tweet is stored in a corresponding column using the Python library openpyxl. For data processing purposes, the pandas data frame was used, while for visualization purposes, Matplotlib was used, both libraries can be installed using the pip package installer for Python.

In addition to using Tweepy library for obtaining relevant tweets relating to milbloggers, two additional datasets were obtained from Kaggle in regard to classifier. The dataset for creating the classifier i.e., training and testing, is based on the Russian Troll Tweets dataset [12] for malicious users while the Sentiment Dataset with 1 Million Tweets is used for non-malicious users [13].

III. TIME ANALYSIS OF TWEETS

The monthly tweet posting interval shown in Fig. 1 demonstrates the increase in the number of posts over the months. A particularly sharp increase was recorded from December to January. The events that took place during the end of 2022 include two successful counter-offensives from the Ukrainian side in the Kharkiv and Kherson regions, in which Ukraine regained a large part of the occupied territory [14]. While on the Russian side during September 21, 2022, the President of the Russian Federation announced partial military mobilization. According to the Defence Minister of the Russian Federation, the approximate number of mobilized persons is at 300,000 [15].

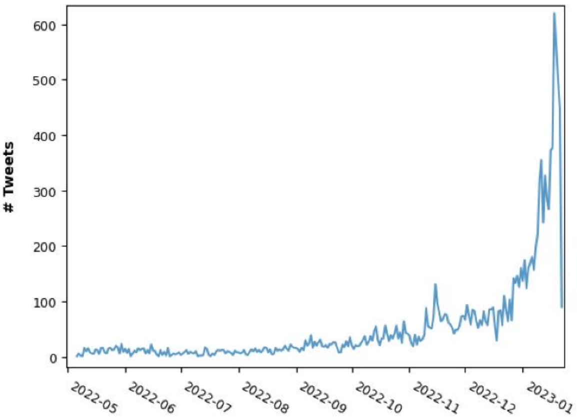

Figure 1. Monthly tweet posting interval

Fig. 2 shows daily interval of posting tweets, with the highest number of posts on Wednesdays and Fridays, and the lowest on Sundays and Mondays. The mentioned distribution is expected since most users who publish tweets are most active during the weekday, and their distribution decreases over the weekend [16].

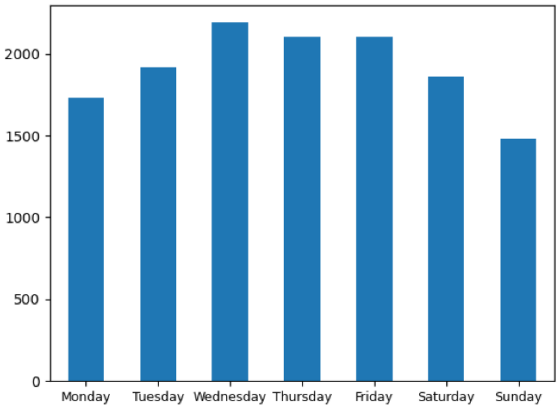

Figure 2. Daily interval of posting tweets

The hourly interval of posting tweets shown in Fig. 3 shows the hourly distribution of tweets where the highest number of tweets was published between 2:00 pm and 6:00 pm, while the lowest number of tweets was published between 3:00 am and 6:00 am.

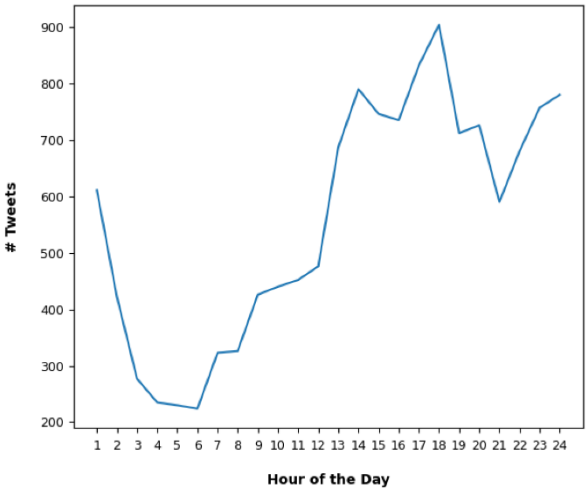

Figure 3. Hourly interval of posting tweets

IV. SENTIMENT ANALYSIS OF TWEETS

Reference [17] states that sentiment classification or sentiment analysis in text classification on social media platform like Twitter is defined as a process of finding out public opinion about an event, product or topic using techniques like machine learning. In it, public opinions are classified into categories like "Positive", "Negative" and "Neutral". Sentiment classification helps organizations to gain insightful knowledge from retrieved data for swift decisions on crucial moments.

Reference [18] points out that sentiment analysis as a field of research, is closely related to (or can be considered a part of) computational linguistics, natural language processing, and text mining. Proceeding from the study of affective state (psychology) and judgment (appraisal theory), this field seeks to answer questions long studied in other areas of discourse using new tools provided by data mining and computational linguistics. Sentiment Analysis has many names. It's often referred to as subjectivity analysis, opinion mining, and appraisal extraction, with some connections to affective computing (computer recognition and expression of emotion).

For the purposes of sentiment analysis, the TextBlob library [19] is used, whose implementation of the SVM classifier used for sentiment analysis shows an accuracy of 88% [20]. TextBlob is a rule-based analysis library that focuses on lexical content and integrates the WordNet corpus for sentiment analysis [21]. In Python the sentiment property of TextBlob returns a named tuple of the form Sentiment(polarity, subjectivity). The polarity score is a float in the range -1.0 and 1.0 while the subjectivity is a float in the range 0.0 and 1.0 where 0.0 is very objective and 1.0 is very subjective. [22]. Tweet polarity histogram shown in Fig. 4 is presenting the polarity of tweets using a visualization line KDE (kernel density estimation), also known as the Parzen's window to estimate the underlying probability density function of a polarity [23].

The visualized polarity in Fig. 4 on the x-axis represents values between -1 and 1, where -1 is the most negative words, and would include words like "disgusting", "distressing" and "miserable", while 1 is the most positive words such as "superb", "greatest" and "magnificent". Apart from the previously mentioned extremes, most of the words do not belong to the mentioned categories, so words like "football", "wood" and "resting" have a neutral value of 0, while words like: "tired", "hard" and "annoyed" have a slightly negative values in contrast to words like: "accomplished", "satisfied" and "gladly" which have slightly positive values.

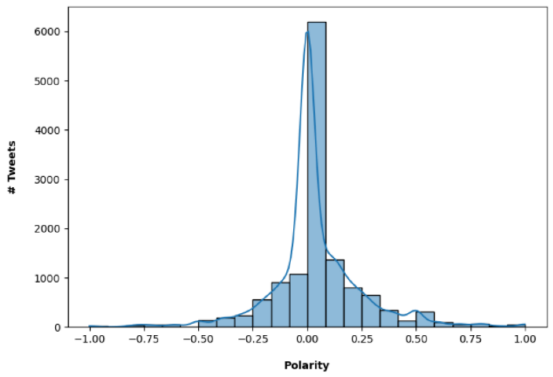

Figure 4. Tweet polarity histogram

That same data is shown in Fig. 5 over time to display correlation between months and polarity associated with each observed month. For more descriptive visualization of time series regarding trends in polarity and subjectivity expanding window and rolling window mean calculations were introduced via calls to pandas.DataFrame.expanding [24] and pandas.DataFrame.rolling [25] methods. For expanding window, number of observations in window is set to value of 1 (default value), while for rolling window, size of the moving window has been set to the time period of 6 hours, both values were calculated by calling:

```
df['sentiment'].expanding(1).mean()
df['sentiment'].rolling('6h').mean()
```

Over time, the change in polarity shows a slight increase, which is evident from the expanding mean which uses more and more observations each time while traversing data. On the other hand, rolling mean uses the same number of observations each time when traversing to define its values. Based on this, the expanding mean value is significantly more flattened than the rolling mean since it takes into account and incorporates new observations that are available. When observing rolling mean we can infer that it is more uniform than the individual representation of each polarity, because it is based on series of averages, while each polarity is raw value without representation of average values.

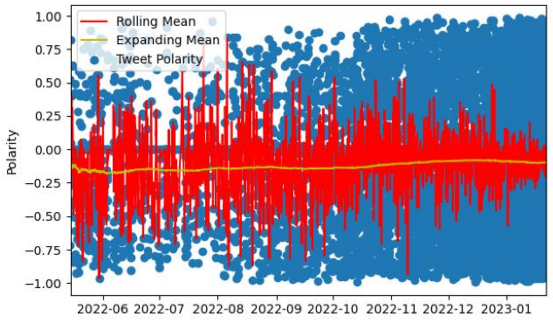

Figure 5. Tweet polarity histogram over time

Fig. 6 shows a histogram of the subjectivity of tweets using the visualization line provided by KDE. Subjectivity is shown on the x-axis with values from 0 to 1, where 0 is very objective, while 1 is very subjective. Thus, words like “great” and “excellent” have a high subjectivity rating, while words like “visit” and “day” have a low subjectivity rating. Low subjectivity indicates that tweets contain more factual information than personal opinions.

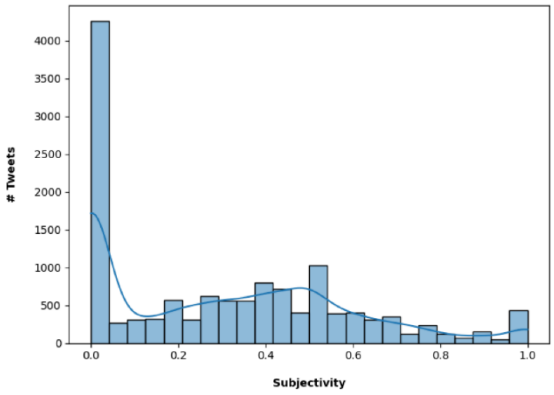

Figure 6. Tweet subjectivity histogram

Fig. 7 is presenting subjectivity over time to display correlation between months and subjectivity associated with each observed month. We can see that over time the subjectivity of tweets increases, which indicates that personal opinions and judgments appear more and more with the passage of time. Increased subjectivity shows a correlation with an increased number of posts, which points to influence operations aimed at triggering readers emotions, therefore each post contains more subjectivity on average.

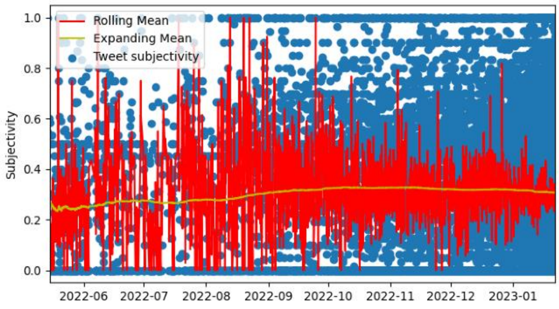

Figure 7. Subjectivity over time

V. SVM MALICIOUS USER CLASSIFICATION

Reference [26] states that SVM (Support Vector Machine) is a supervised machine learning algorithm that performs binary and multiway classification (pattern recognition) of the data into user defined categories. Support Vector Machines maps non-linearly separable training vectors in input space to linearly separable higher dimensional feature space and finds a separating hyper plane with maximal margin in that higher dimensional space. SVM is generally used for text categorization while also achieving good performance in high-dimensional feature space [27].

SVC (Support Vector Classification) implementation based on libsvm was used to classify malicious users. The fit time scales at least quadratically with the number of samples and may be impractical beyond tens of thousands of samples [28] with this limitation in mind, a limit of 20 thousand tweets was imposed for input into the classification model.

From each individual dataset used to create classifier 10,000 tweets were randomly selected, tagging each corresponding tweet with appropriate label, where 1 was used if it belongs to the Russian Troll Tweets dataset, while 0 was used for tweets belonging to Sentiment Dataset with 1 Million Tweets dataset. An additional filter on the Sentiment Dataset with 1 Million Tweets dataset was introduced where the language for selected tweets is specified as English since dataset also contains tweets in other languages such as: French, Spanish, Portuguese and Japanese.

Fig. 8 demonstrates steps that are involved in SVM classifier construction after tweets were appropriately labeled. These steps are separated in three categories: a) data preprocessing, b) model training and c) classification.

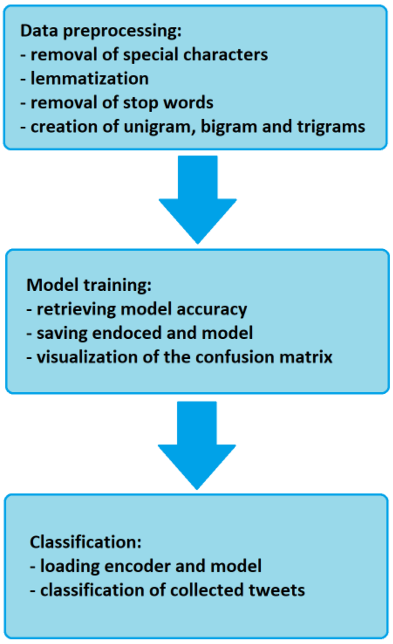
Data preprocessing:
- removal of special characters
- lemmatization
- removal of stop words
- creation of unigram, bigram and trigrams
Model training:
- retrieving model accuracy
- saving endoced and model
- visualization of the confusion matrix
Classification:
- loading encoder and model
- classification of collected tweets

Figure 8. Visualization of SVM classifier construction

Data preprocessing is done in 4 separate steps:

a) Removal of special characters: is done by using regex r'[^a-zA-Z0-9\s]' to remove any special characters and to keep only numbers and alphabet letters inside each tweet.

b) Lemmatization: is achieved by implementation of lemmatization words using NLTK which provides WordNetLemmatizer class which is a slim cover wrapped around the wordnetCorpus. This class makes use of a function called Morphy() to the class to find a root word/lemma [29].

c) Removal of stop words: is done by providing parameter of stop_words to TfidfVectorizer class constructor with value of “english".

d) Creation of unigram, bigram and trigrams: are also provided as a constructor parameter to TfidfVectorizer ngram_range whose value is tuple (1,3).

By using TfidfVectorizer sklearn class collection tweets

TABLE III. OVERVIEW OF CLASSIFIED 13,381 TWEETS

Total	Malicous	Nonmalicous
13,381	3,586	9,795

a. Velika slova

were converted to a matrix of TF-IDF (term frequency–inverse document frequency) features together with removal of stop words while individual words were converted to ngrams. TF-IDF uses the frequency of words to determine how relevant those words are to a given document. The goal of TfidfVectorizer sklearn class is to give higher weightings to terms that appear often in a particular document, but not in many documents. If a word appears often in many of the documents, it is not a good feature for discriminating between classes. Likewise, if a word appears often in some documents and not in others, it is likely a good word for discriminating between classes [30].

TfidfVectorizer when used on 10,000 tweets, that were previously randomly selected and labeled with 1 (tweeets from Russian Troll Tweets dataset), produced vocabulary contains 143,698 words. Words such as: "rt", "trump", "http", "make", "just", "right", "sad", "need", "life", "politics"... showed high TF-IDF scores compared to other words inside this vocabulary.

Train and test split was made with respective 80:20 split, with train dataset containing 16,000 tweets, while test dataset contains 4,000 tweets. TfidfVectorizer produced vocabulary of 238,223 words when it was used on training dataset and 65,112 words when used on test dataset. Confusion matrix shown in Fig. 9 is used to describe the classification summary in a table. It can be interpreted as the summary of the predicted and actual data. As a result, for confusion matrix values tags were provided for: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) with their corresponding values [31]. Based on given values, an accuracy score of 93.95% is achieved, which is obtained by dividing the sum of TN and TP with the size of the test dataset. We can also calculate F1 score by using formula F1 = 2TP / (2TP + FP + FN), that will give us value of 0.9424.

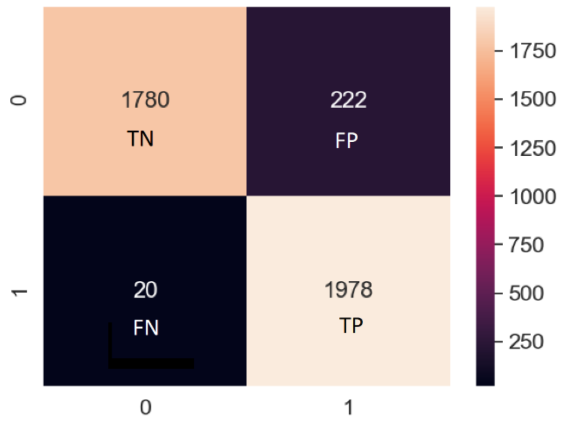

Figure 9. Confusion matrix based on SVC using TfidfVectorizer

In context of this paper malicious tweets are defined as a tweets that are part of Russian Troll Tweets dataset and any new tweets that are classified as such based on training of this dataset. By running the classifier on previously collected tweets, a classification is obtained with 1 for malicious and 0 for nonmalicious tweets, where 26.8% of tweets are classified as malicious tweets.

Creating a histogram Fig. 10 on the obtained data of the SVM classifier, it is possible to visualize the trend of the of malicious tweets, which become less and less malicious with the passage of time.

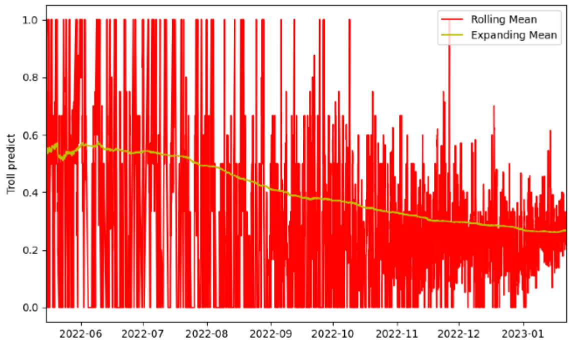

Figure 10. Malicious tweets classification over time

The movement of the trend where the classification of individual tweets as malicious is decreasing shows a clear focus on a controlled campaign that aims to avoid the classification of tweet as malicious with the change of tweet composition from previous campaigns on which the classifier was trained. Such behavior would be expected since the development of the model depends on previously classified malicious users, and by changing their behavior, that is, their posts, it is much more difficult to identify them without the existence of previously published datasets that contains classified malicious users.

VI. CONCLUSION

The Russian invasion of Ukraine on February 24, 2022 triggered a conflict of narratives and influence operations aimed at directing the behavior of readers towards pro-Russian sentiment through the intensity of posts. The growth in the number of tweets covered by pro-Russian milbloggers, especially at the end of the 2022, is related to the Kherson counteroffensive and the Kharkiv counteroffensive, during which a significant area that was previously under Russian occupation was liberated.

Aforementioned events weakened the credibility of the Russian military leadership in the decisions that were made, which affected the public opinion towards the war in Ukraine, as well as the morale of the Russian troops involved in the conflict. Both events caused a more significant start of Russian attacks on critical infrastructure that correspond to the narrative and preparation of Russia for a long war, and thus the preparation of the public through influence operations.

In addition to the increase in the volume of tweets as a stage of preparation for the narrative via Twitter, the approach to the very sentiment of published tweets is also changing, showing a slight increase in subjectivity and increasing polarity with the passage of time. In addition to the growth of tweets, subjectivity and polarity, the classification of malicious users is in a more significant decline than the previously mentioned growth of other parameters. Which indicates that the new tweets are written much more carefully for the purpose of influence operations, and as such show a lower degree of detection, where the propaganda aspect of the tweets is reduced, but not completely removed which happens through an increased volume that has the possibility of reaching a larger number of users.

Based on this work, numerous possibilities for future work are opened up by monitoring the trend of posting of current users, as well as increasing the pool of Twitter profiles from which tweets are collected. Such research can also encompass sentiment analysis as well as LDA topic modeling to see if detected malicious users have influence on Ukraine, United States of America and European Union political process by changing and shaping their policies in regard to influence operations. This can be done by collecting tweets from members of Ukraine parliament (Verkhovna Rada of Ukraine), United States Congress as well as European Parliament members that in the same time period as influence operations were ongoing.

REFERENCES

[1] Haq E. U, Tyson G, Lee L. H, Braud T, Hui P. Twitter dataset for 2022 russo-ukrainian crisis, 2022, pp. 1-2.

[2] Ślufińska, M. "The Russia-Ukraine War", Information Security Policy, 2022, pp. 69-70.

[3] Nye J. S. “Soft power: the evolution of a concept”, Journal of Political Power, 2021, 14, pp. 202-203.

[4] Clark M, Barros G, Stepanenko K. “Russian Offensive Campaign Assessment”, Institute for the Study of War, 2022

[5] Hoskins A, Shchelin P. “The War Feed: Digital War in Plain Sight”, American Behavioral Scientist, 2022

[6] Kagan, F. W., Barros, G., & Stepanenko, K. “Russian offensive campaign assessment”, Institute for the Study of War, 2022, pp. 2-3.

[7] Rybar, available at: https://t.me/rybar, [accessed: 01.15.2023].

[8] Rybar in English, available at: https://twitter.com/rybar_en [accessed: 01.15.2023].

[9] Unmasking Russia’s influential pro-war ‘Rybar’ Telegram channel, available at: https://thebell.io/en/unmasking-russia-s-influential-pro-war-rybar-telegram-channel/ [accessed: 01.20.2023].

[10] The Bell, available at: https://twitter.com/thebell_io/status/1595106311133597696 [accessed: 01.20.2023].

[11] Tsakiris A. The Portrayal of the 2022 Russian Invasion of Ukraine onSocial Media, 2022, pp. 20.

[12] FIVETHIRTYEIGHT, Russian Troll Tweets, available at: https://www.kaggle.com/datasets/fivethirtyeight/russian-troll-tweets [accessed: 01.27.2023].

[13] TARIQ M., Sentiment Dataset with 1 Million Tweets, available at: https://www.kaggle.com/datasets/tariqsays/sentiment-dataset-with-1-million-tweets [accessed: 01.27.2023].

[14] CHARAP, S. and PRIEBE, M. Avoiding a Long War, 2023, pp. 6.

[15] Zmyvalova E. “The Rights of Indigenous Peoples of Russia after Partial Military Mobilization”, Arctic Review on Law and Politics, 2023, 14, pp. 71.

[16] Honey C. and Herring S. C. “Beyond microblogging: Conversation and collaboration via Twitter”, Hawaii International Conference on System Sciences, 2009, 42, pp. 2.

[17] Shekhawat B. S. Sentiment Classification of Current Public Opinion on BREXIT: Naïve Bayes Classifier Model vs Python’s TextBlob Approach (Doctoral dissertation, Dublin, National College of Ireland), 2019, pp. 2.

[18] Mejova Y. “Sentiment analysis: An overview”, University of Iowa, Computer Science Department, 2009, pp. 5.

[19] TextBlob: Simplified Text Processing, available at: https://textblob.readthedocs.io/en/dev/ [accessed: 01.25.2023].

[20] Falco X, Witten R, Zhou R. SENTIMENT AND OBJECTIVITY CLASSIFICATION, 2009, pp. 23.

[21] Madhu Fadhli I, Hlaoua L, Omri M. N. "Sentiment analysis CSAM model to discover pertinent conversations in twitter microblogs", I. J. Computer Network and Information Security, 2022, 5, pp. 34.

[22] Madhu S. “An approach to analyze suicidal tendency in blogs and tweets using Sentiment Analysis”, I. J. Scientific Research in Computer Science and Engineering, 2018, 6, pp. 36.

[23] Chen, Y.C. “A tutorial on kernel density estimation and recent advances”, Biostatistics & Epidemiology, 2017, 1, pp. 162.

[24] pandas.DataFrame.expanding, available at: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.expanding.html [accessed: 03.14.2023].

[25] pandas.DataFrame.rolling, available at: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rolling.html [accessed: 03.14.2023].

[26] Polavarapu N, Navathe S. B, Ramnarayanan R, Haque A, Sahay S, Liu Y. “Investigation into biomedical literature classification using support vector machines”, Computational Systems Bioinformatics Conference, 2005, pp. 368.

[27] Bholane S. D. and Gore, D. “Sentiment analysis on twitter data using support vector machine”, International Journal of Computer Science Trends and Technology (IJCST), 2016, 4, pp. 368.

[28] C-Support Vector Classification, available at:

https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html [accessed: 01.27.2023].

[29] Khyani D, Siddhartha B. S, Niveditha N. M, Divya B. M. “An interpretation of lemmatization and stemming in natural language processing”, Journal of University of Shanghai for Science and Technology 2021, 22, pp. 355.

[30] Djuve K. and Burris, J. W. “A case study on the dialect identification of twitter tweets using natural language processing and machine learning”, Journal of Computing Sciences in Colleges, 2019, 34, pp.67.

[31] Adiba F. I, Islam T, Kaiser M. S, Mahmud M. and Rahman M. A. “Effect of corpora on classification of fake news using naive Bayes classifier”. International Journal of Automation, Artificial Intelligence and Machine Learning, 2020, 1, pp. 88.