Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 09, No. 03, September 2023.

PREDICTION OF PEOPLE SENTIMENTS ON TWITTER USING MACHINE LEARNING CLASSIFIERS DURING RUSSIAN AGGRESSION IN UKRAINE

Mohammed Rashad Baker[1*], Yalmaz Najmaldeen Taher [2] and Kamal H. Jihad [3]

(Received: 12-Feb.-2023, Revised: 29-Apr.-2023, Accepted: 18-May-2023)

ABSTRACT

Social media has become an excellent way to discover people's thoughts about various topics and situations. In recent years, many studies have focused on social media during crises, including natural disasters or wars caused by individuals. This study examines how people expressed their feelings on Twitter during the Russian aggression on Ukraine. This study met two goals: the collected data was unique and it used Machine Learning (ML) to classify the tweets based on their effect on people's feelings. The first goal was to find the most relevant hashtags about aggression to locate the dataset. The second goal was to use several well-known ML models to organize the tweets into groups. The experimental results have shown that most of the performed ML classifiers have higher accuracy with a balanced dataset. However, the findings of the demonstrated experiments using data-balancing strategies would not necessarily indicate that all classes would perform better. Therefore, it is essential to highlight the importance of comparing and contrasting the data-balancing strategies employed in Sentiment Analysis (SA) and ML studies, including more classifiers and a more comprehensive range of use cases.

KEYWORDS

Sentiment analysis, Machine learning, Classification algorithm, Imbalanced data classification, Russian aggression in Ukraine.

1. INTRODUCTION

Crises have a major impact on human societies, altering the lives of individuals in significant ways. To understand the reactions of societies in times of crises, it is crucial to listen to people's ideas and comprehend their sentiments. Therefore, Sentiment Analysis (SA) has emerged as a vital subject of study in Natural Language Processing (NLP) and information extraction [1]. It seeks to evaluate a wide range of information, eliciting writers' emotions reflected in positive or negative words [2]. With the rise of social networking platforms, such as Facebook, Twitter, Linkedin and others, people have gained significant power in expressing and exchanging opinions about political or social events and inevitable social crises [3]. Nevertheless, understanding people's behaviors becomes challenging during crises because of the sheer volume of instructive messages, emotional outbursts, helpful safety suggestions and rumors. It is essential to leverage SA to better manage and regulate a crisis [4].

Natural disasters pose a significant challenge for societies and real-time sentiment analysis on social-media platforms (such as Twitter) can play a crucial role in saving lives [5]-[6]. Twitter's micro-blogging service allows users to share messages about events and news worldwide, using hashtags to follow hot topics. In the case of natural disasters, SA can be used to analyze tweets related to events, like the California Campfires (considered one of the most damaging and destructive wildfires in the history of California) [7]. However, there is a lack of research on the SA of natural disasters, causing negative impacts on society in many respects. More research attention and efforts need to study people's reactions to disasters. The studies must include mitigating, preparing, recovering and responding to disasters while reducing damage to citizens and economies [8]. During conflicts or aggressions caused by natural or human factors, opinions and sentiments can be expressed using social-media platforms, like Twitter [9].

Real-time assessment of public opinion expressed in tweets can aid authorities in developing early, response strategies. For example, a study examined the rule of the Taliban in Afghanistan after the

1. M. R. Baker is with Software Department, College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq. Email: mohammed.rashad@uokirkuk.edu.iq
2. Y. N. Taher is with College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq. Email: yalmaz.science@uokirkuk.edu.iq
3. K. H. Jihad is with College of Science, University of Kirkuk, Kirkuk, Iraq. Email: kamal.jihad@uokirkuk.edu.iq

withdrawal of US soldiers, using public opinion expressed in tweets [10]. Additionally, a new method was proposed for real-time sentiment analysis on the current Refugee Crisis to provide some prediction on polarity types of political improvement based on Twitter data [11]. Machine Learning (ML) techniques were used to present a method for sentiment analysis on Twitter data, comprising tweets about Afghanistan. The study identified algorithms and measures for evaluating the performance of supervised ML classifiers on tweets on the US war in Afghanistan [9]. The research focused on the refugee crisis. Accordingly, through binomial classification of positive and negative, ML algorithms were applied to obtain final-level decisions regarding the number of individuals commenting in support of refugees [11]. It is clear that exploring feelings has become critical, particularly in studying and processing natural languages [12]. More research is needed to leverage sentiment analysis to understand people's reactions during crises and develop strategies for effective crisis management.

The Russian aggression on Ukraine has raised questions about the formation and evolution of group identities during times of political tension [13]. Existing research suggests that insecurity, competition over resources and threat perception from out-groups increase ethnic-identity salience. However, Metzger et al. [14] proposed a novel approach using Ukrainian Twitter users' language preferences to examine this issue. The study found that key political events during the Ukrainian crisis did not lead to a reversion to language preferences, but following the annexation of Crimea, both Russian and Ukrainian Twitter users began using Russian tweets with greater frequency. Driscoll and Steinert-Threlkeld [15] suggested that social media provides insight into political attitudes and the study mapped the evolution of Russian-speaking communities' attitudes towards the conflict. The results show that the Russian-Ukrainian interstate border moved as far as the Russian military could advance without incurring occupation costs. These studies offer insights into the complex relationship between language, identity and political conflict and provide a basis for future research [14]-[15].

However, there has been comparatively less research into the social and emotional aspects of the Russian aggression in Ukraine, particularly the sentiments of those who are affected by it [16]. Understanding the sentiments of individuals towards aggression can provide several benefits. Firstly, it can provide insights into the effectiveness of various propaganda and messaging campaigns used by the parties involved, which can be useful in designing more effective messaging strategies. Secondly, it can provide insights into the emotional impact of aggression on individuals, which can help researchers better understand the psychological toll of aggression and identify areas where emotional support may be needed. Thirdly, sentiment analysis can identify potential sources of tension and aggression escalation. By monitoring sentiment trends over time, researchers can identify periods of heightened tension and negative sentiments that can signal the potential for further aggression or violence. Sentiment analysis of the Russian aggression in Ukraine is essential, as it can provide valuable insights into individuals' emotions, opinions and attitudes towards the aggression. It can be used to design more effective messaging strategies, provide emotional support to those affected by the aggression and identify potential sources of tension and aggression escalation.

ML techniques are important for sentiment analysis of the Russian aggression in Ukraine, because they enable the processing of large amounts of data quickly and accurately, allowing for a more comprehensive analysis of the sentiments expressed by individuals affected by the aggression. Additionally, ML algorithms can learn from the patterns and characteristics of the sentiment data to improve the accuracy of the sentiment analysis. This can provide valuable insights into individuals' emotions, opinions and attitudes toward aggression and can help identify potential sources of tension and aggression escalation. Moreover, ML techniques can be used to evaluate the effectiveness of different propaganda and messaging campaigns used by the parties involved, which can inform the design of more effective messaging strategies.

This study aims to predict people's sentiments on Twitter during Russian aggression [13] in Ukraine by using machine-learning classifiers while investigating how individuals respond and behave during a crisis, particularly in the context of a war or aggression. In this study, we aim to examine the potential of utilizing sentiment analysis using machine-learning techniques to comprehend community behavior and attitudes toward the war. It includes the degree to which the ML model can assist in comprehending community behavior, the level of correspondence between the observations and the actual user sentiments analyzed from the tweets, as well as the extent to which sentiments are uniform within and across regions. The major contributions of this study comprise the following:

1. A machine learning-based sentiment detection model for Twitter feeds concerning the war.
2. Using several machine-learning models for classifying sentiment polarity and emotions.
3. Intriguing insights into collective reactions to the war on social media could aid in informing decision-making processes and potentially contribute to developing more accurate and effective sentiment-analysis tools.

The structure of this article is as follows. Section 2 introduces the works related to this topic. In Section 3, we present our methodology. Section 4 discusses the experimental results. The conclusion will be drawn in Section 5.

2. RELATED WORKS

This section of the paper is organized into two sub-sections. The first sub-section provides a comprehensive literature review on sentiment analysis. The second sub-section focuses on related works on Russian aggression in Ukraine and how sentiment analysis has been used in this context.

2.1 Sentiment Analysis Related Works

SA has garnered significant interest in recent years due to the prominence of social-media sites, such as Twitter and Facebook. In addition, the availability of voluminous data in tweets, reviews and comments expedited its development. As a result, there is a substantial body of literature on SA [10]. The proposed method detects fake news using sentiments with positive and negative scores. Elmurngi and Gherbi [17] used statistical approaches to assess the efficacy of spambot systems in the SA arena. The task-specific precision of various ML models is tested. Wael et al. [18] used SA to identify Western media's bias in the Palestinian-Israeli crisis. This process includes finding deceptive terms, vocabularies and idioms used to sway public opinion about the Israel-Palestine problem.

The refugee issue was also considered utilizing SA. For instance, Ozturk and Ayvaz [19] analyzed Turkish and English tweets to address the challenges of Syrian refugees. They examined public feelings and opinions regarding the Syrian refugee situation. The results demonstrated a substantial variation in sentiments between Turkish and English tweets. The data also indicated that Turkish tweets include more optimistic sentiments. A comparative examination revealed that Turkish tweets contain more positive than negative or neutral sentiments toward Syrian refugees. Another considered issue was terrorism; for instance, Mansour [20] conducted SA on tweets related to ISIS to gain insights into how people feel about acts related to terrorism. The Term Frequency-Inverse Document Frequency (TF-IDF) approach was applied in the study to perform SA on tweets on ISIS.

Other researchers have used Twitter SA to determine various public opinions and feelings expressed during crises, such as civil wars and natural disasters [21]-[22]. Identifying these feelings is important for understanding the situations' dynamics and their emotional impact on affected people. A study shows that debriefing during a disaster can help authorities develop critical situational awareness and other programs to manage future events [23]. Studies showed that users' emotions fluctuate depending on location and proximity to the disaster site. For example, a study assessed the situation and public opinion regarding Brexit, in which more than 16 million tweets were collected. This study uncovered the most popular daily Twitter debates and discovered a positive correlation between Twitter's attitude towards Brexit and the British-pound exchange rate using the VADER library [24].

Protests have become more common in recent years and researchers are interested in understanding the emotions and sentiments expressed during these events through social media. Field et al. [25] used natural-language processing techniques to analyze emotions in tweets about the 2020 Black Lives Matter protests. They found that positive emotions, such as pride and hope, were prevalent in tweets with pro-BlackLivesMatter hashtags, contradicting stereotypical portrayals of protesters as perpetuating anger and outrage. Won et al. [26] developed a visual model that uses convolutional neural networks to classify the presence of protesters in an image and predict their visual attributes, perceived violence and exhibited emotions. They also released a novel dataset of 40,764 protest images with various annotations of visual attributes and sentiments. Steinert-Threlkeld and Joo [27] introduced the Multimodal Chile & Venezuela Protest Event Dataset (MMCHIVED), which contains city-day event data using a new source of data, text and images shared on social media, enabling the improved measurement of variables, such as protest size, protester and state violence, protesters' demographics and their emotions. Overall, these studies demonstrate the value of analyzing social-media data to understand the emotions and sentiments expressed during protests.

According to the information reported above, SA has become a significant research topic in artificial intelligence. According to a survey of the available literature, various Twitter SA research employs classic ML algorithms to estimate sentiments from tweets [28]–[30]. These approaches tackle SA problems as if they were text-classification problems. These algorithms treat SA problems as text-classification problems and have been found to provide high accuracy with fewer computational resources. The classic ML algorithms commonly used for SA of Twitter data include Naive Bayes (NB), Random Forest (RF) and Logistic Regression (LR), among others. These algorithms provide strong accuracies with fewer computer resources and are used widely in SA of Twitter data [29].

In this work, we use a pre-annotated dataset using RoBERTa and TextBlob. Next, we apply various ML classifiers for SA to analyze tweets about the Russian-Ukrainian armed conflict. To the best of our knowledge, this research is among a few that seek to provide valuable insights into tweet content related to the Russian-Ukrainian war. Findings can be a reputable source of information to help governments and international organizations understand social-media trends and public views on the situation in Ukraine.

2.2 Russian Aggression in Ukraine Related works

Numerous studies have investigated the Russian aggression in Ukraine's online social networks (OSNs), focusing on Twitter and Reddit data to uncover hidden insights, disinformation campaigns and abnormal patterns [31]. There is a growing need for further research in the field, including aspect-based sentiment analysis (ABSA), to mine and analyze large datasets on OSNs. Hanley et al. [32] found differences in news coverage among Western, Russian and Chinese press outlets, with Russian media focusing on the purported justifications for the military operation and Chinese news media concentrating on the aggression's diplomatic and economic consequences. A novel lexicon-based unsupervised sentiment-analysis method was proposed by Guerra et al. [33] to measure "hope" and "fear" using Reddit.com as the main source of human reactions to daily events during nearly the first three months of the aggression.

Propaganda and misinformation were studied by Pierri et al. [34] on Facebook and Twitter during the first few months of the Russian aggression in Ukraine. They found that superspreaders played a disproportionate role in amplifying unreliable content and the political leaning of Facebook pages and Twitter users sharing propaganda was more right-leaning than average. In another study, Agarwal et al. [35] analyzed the emotional sentiments of tweets acquired during the peak war period, from December 31, 2021 to March 03, 2022. The study found more negative tweets than positive ones. It provided insights into the spread and influence of different categories of tweets, highlighting the need for further research on dynamic sentiment analysis.

In the study by Vyas et al. [36], a framework was developed to automatically classify distinct societal emotions related to the Russia-Ukraine War (RUW) on Twitter. The authors found that most tweets describe the RUW in key terms related more to Ukraine than to Russia and that 81% of Twitter users surveyed showed a neutral position toward the aggression. In another study, Vyas et al. [31] proposed a hybrid framework to automatically extract positive, negative and neutral sentiments from tweets related to the COVID-19 pandemic and classify them through machine-learning techniques.

Ibar-Alonso et al. [37] conducted a social-listening analysis on Twitter to assess sentiments and emotions regarding green energy during the onset of the 2022 Russian aggression in Ukraine. They found that the aggression changed society's sentiments about an energy transition to green energy, with negative feelings and emotions emerging in green-energy tweeters once the aggression started. The emotion of confidence increased as the aggression drove all countries to promote a rapid transition to greener-energy sources.

In the study by Chen et al. [38], the authors analyzed the public opinion warfare related to the Rural-Urban Waiver (RUW) in Chinese Weibo texts. They used Latent Dirichlet Allocation for unsupervised clustering and an opinion adversarial evolution algorithm to dynamically model the dominant degree of an opinion in the evolutionary processes. The authors released a dataset of Chinese Weibo associated with the RUW and proposed a data-driven approach for analyzing opinion warfare in cyber-physical-social systems. The study calls for further expansion of data collection and analysis from multiple perspectives and the design of unsupervised clustering methods for complex social texts to improve opinion recognition.

Garcia and Cunanan-Yabut [39] analyzed the sentiments and emotions of the international community towards the Russian invasion of Ukraine using tweets posted on the first day in the #UkraineRussia hashtag. The results showed that negative sentiments were more prevalent and sadness was the most salient emotion. The study highlights the potential of social media, particularly Twitter, as a vehicle for mass communication that governments and politicians can use as a source of public opinion. Future research could continue examining the platform as a channel for public participation in peacemaking.

Benjamin Džubur et al. [40] combined sentiment and network analysis approaches to produce various insights into the discussion of the Russian aggression on Ukraine in their study. They discovered that most users support Ukraine and that the most critical accounts belong to political leaders, as well as relevant organizations or media outlets that actively report on the aggression. Apart from a few pro-Russia communities, all the groups express support for Ukraine to some degree. The study suggests that future research should focus on more thoughtful data collection and thorough analysis of various aspects of the networks.

3. METHODOLOGY

This section covers many methodological approaches that were taken throughout this research. Figure 1 shows that our methodology comprises four main steps. The first step is to examine the data-collection process and identify keywords. In the following step, pre-processing procedures have been applied to the dataset, starting with initial filtering and continuing to complete processing. After concluding the previous step, the next step addresses the topic of preparing the dataset for ML classification with a discussion that includes activities, such as the annotation technique and feature extraction. The last step applies the ML prediction models suggested for this research and discusses the results of the model performance.

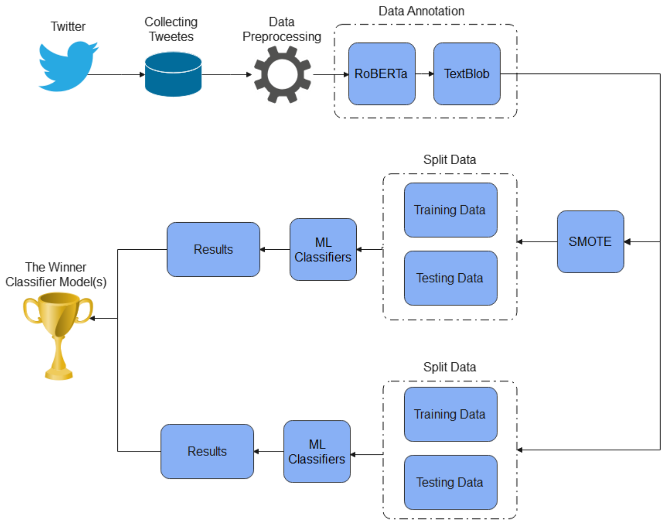

Figure 1. The steps of our methodology.

3.1 Data Collection

This research aims to comprehensively and representatively collect tweets about the aggression and crisis between Russia and Ukraine on Twitter. Prior research on aggression and crisis was conducted to identify the most relevant and popular data-collection hashtags. This involved an extensive review of news reports, social-media discussions and other relevant sources in identifying the key themes, topics and issues related to aggression.

Data collection for tweets related to aggression and crisis started on February 24, 2022 and lasted until July 4, 2022, on Twitter, which is the most popular platform for expressing thoughts and opinions [41]. The selected hashtags included "#UkraineVSRussia, #UkraineConflict, #UkraineCrisis, #UkraineWarCrimes, #stopwar, #UkraineWar, #UkraineRussia and #ukrainianwar," based on prior research and their popularity. The streaming Twitter API justifies its use, since it can give every tweet for a given set of search terms [41]. However, Twint, a popular third-party tool, is also favored by researchers because of its flexibility and ability to scrape tweets without limitations or barriers [1]. We used the Twint project tool to scrape 1,245,791 tweets using these hashtags. We deliberately chose to use Twint over the Twitter API due to its ability to access all publicly available tweets, enabling historical-data retrieval that was crucial for our study. Twint also has advantages over the Twitter API, such as collecting data without authentication, avoiding rate limits and enabling advanced search queries. All tweets collected using Twint with these hashtags were saved as comma-separated values (.csv) for analysis in this research. To promote transparency and facilitate future research, we have made the dataset used in this manuscript publicly available to other researchers by uploading it to Mendeley [42].

3.2 Pre-processing

While tweets in other languages can provide insights into public sentiments during the ongoing crisis in Ukraine, our study focuses solely on English tweets. English is the most widely used language on Twitter, which ensured a substantial sample size for analysis. Furthermore, gathering and translating tweets in other languages would have required additional language processing, which was outside the scope of our study. Our dataset includes crisis-related keywords and was not restricted by location, resulting in tweets from all over the world. Before using Twitter's data for further processing, the data must undergo a filtering process to eliminate things with no relation. Several Python tools and libraries were used. The first step is to normalize the data to clean it up. As part of this process, non-English tweets, emojis, URLs and other special characters are filtered out. As a part of normalization, all capital letters are turned into small letters to simplify using text in the analysis [43]. The following step is tokenization. Tokenization is a pre-processing method that can divide any text into small pieces called "tokens." All special characters are taken out during tokenization, the spaces between words are found and abbreviations, numbers and other special characters are processed. The last step is removing English-language stop words like "is," "a," and "an," which do not add much to the text about a specific topic.

Figure 2 shows the distribution of the weekly tweets. The analysis indicates that the number of tweets per week varies significantly, ranging from a minimum of 677 to a maximum of 88778, with an average of 18112.3 tweets per week. This indicates that there is a lot of fluctuation in the amount of Twitter activity, which can be influenced by various factors, such as current events, popular topics and the behavior of users. Analyzing this data can help individuals and organizations better understand trends and patterns in Twitter activity and use this information to inform their social-media strategy.

Figure 3 displays a word cloud of the collected dataset related to the topic of Ukraine and Russia's conflict, with words, such as "destroy," "kill," "war," "force" and "military," alongside popular words, like "Ukraine," "Russia" and "War," suggesting a negative sentiment towards the conflict. The inclusion of words, like "peace," "help," "make," "child" and "live" in the word cloud could indicate a shift in sentiments towards a more positive direction. These words suggest the possibility of a desire for a resolution to the conflict and a focus on the impact of the conflict on civilians, particularly children and families.

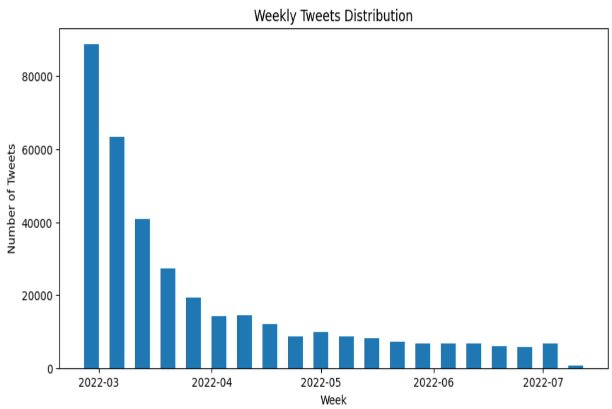

Figure 2. Weekly tweets distribution.

Figure 3. The word cloud generated from collected dataset.

Table 1 shows a list of 10 random tweets and their polarity scores generated after applying the RoBERTa model to perform sentiment analysis. The polarity scores indicate the sentiment expressed in the tweets, where a positive score denotes a positive sentiment and a negative score denotes a negative sentiment.

Table 1. List of 10 random tweets along with their polarity scores.

No.	Time	Tweet	Polarity
1	2022-02-25 15:50:54	foreign policy morality individuals nations act interest principles ordinary citizens force patriotism moral individuals spar nations moral scrutiny	-0.093
2	2022-02-26 22:13:04	miss Ukraine also join Ukrainian army let forget women fight beloved country	0.703
3	2022-02-27 07:26:20	watch Ukrainian news outlets make cry understand evil understand children die war	-1.02
4	2022-02-27 14:00:29	man find mine near Berdyansk pick hand cigarette mouth move away woods	0.101
5	2022-03-03 14:17:11	word speak furiously become cause unrest life war become cause destruction many generations	0.503
6	2022-03-08 02:05:02	deputy state Sherman say may become harder come days	-0.101
7	2022-03-10 15:05:21	experts keep say cannot faceoff Putin military Putin military commit war crimes already say cannot engage stink chamberlain criticism NATO us need make clear Russia	-0.033
8	2022-03-24 14:00:07	talk love ones Ukraine distress news via situation Ukraine crucial importance talk love ones supportive sensitive way	0.320
9	2022-04-30 02:44:43	Ukrainian girls consider one beautiful world today also defend country brave strong courageous different one thing unite desire win	0.576
10	2022-06-23 22:33:37	Ukraine receive long range rocket system Russian official threaten strike us embassy Kyiv	-0.025

A score of zero indicates a neutral sentiment. The table shows that the tweets cover various Ukraine-related topics, including foreign policy, military, news outlets, war crimes and personal relationships. The polarity scores of the tweets generated by RoBERTa and TextBlob vary slightly, indicating that different models may produce slightly different results depending on the text being analyzed. The tweet with the highest polarity score is number 2, expressing a positive sentiment towards Ukraine and its women who fight for their country. The tweet with the lowest polarity score is number 3, which expresses a highly negative sentiment towards Ukrainian news outlets and the reality of war. The other tweets have polarity scores that fall in between these two extremes, with some expressing positive sentiments (tweets 4, 5, 8 and 9), some expressing negative sentiments (tweets 1, 3, 6 and 10) and one tweet with neutral sentiment (tweet 7). Overall, the table provides a glimpse into the sentiments expressed on Twitter towards Ukraine during the time period covered by the tweets. The results demonstrate the usefulness of sentiment analysis in understanding public opinion and highlighting the importance of choosing the appropriate sentiment-analysis model for the specific context and purpose of the analysis.

3.3 Data Annotation

According to the research’s main case for classification, the tweet annotation labels are assigned to various categories [44]. In our study, we fine-tuned the pre-trained Roberta model on a smaller annotated dataset relevant to our research domain. The annotated dataset consisted of text documents with labeled sentiment scores ranging from negative to positive. We used this dataset to train the model to identify and classify sentiments in the text data. Our study used the state-of-the-art RoBERTa model based on transformer architecture to perform sentiment analysis on text data for polarity. This allowed the model to learn contextual representations of words and sentences that could be fine-tuned for specific natural-language processing tasks. We fine-tuned the pre-trained Roberta model on a smaller annotated dataset consisting of text documents with labeled sentiment scores ranging from negative to positive, using the Hugging Face Transformers library to implement the PyTorch implementation of the model. To define the polarity, we used TextBlob on the result that we obtained from the RoBERTa transformer. During the model’s training, we used a batch size of 32 and trained it for 10 epochs with a learning rate of 2e-5 while also using early stopping to prevent overfitting. Accordingly, the dataset was annotated into two categories for binary classification using TextBlob. The two categories are; negative tweets regarding the aggression annotated by 0 and positive tweets regarding the aggression annotated by 1. A total of 362,246 relevant tweets were considered after applying labeling to our collected dataset. The number of tweets in each of the two categories is depicted in Figure 4.

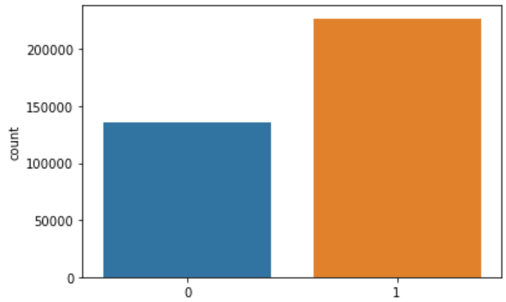

Figure 4. Distribution of categories in the dataset.

As we can notice from the figure, there are two classes of repented tweets. The first class, labeled 1, has 226,936 tweets about war and the second class, labeled 0, has 135,310 tweets about the same conversations during the aggression. The number of tweets in the first category is greater than in the second category, which could affect the results of the ML classifiers in the studied case. Synthetic Minority Oversampling Technique (SMOTE), a distribution-balancing technique, will be implemented to address this problem. Furthermore, classification findings will be shown before and after applying the SMOTE to establish how this technique improves categorization.

3.4 SMOTE

SMOTE is used in ML research for data balancing. Generating data samples of minority classification labels, such as the number of samples from each group, is nearly equal [45]. As described before, the dataset did not provide an equal distribution of categories, which can cause the ML models to overfit. To address the class imbalance in our dataset, we opted for the SMOTE, as it provides several benefits over other approaches. SMOTE is widely used in text classification and has proven effective in dealing with imbalanced datasets. One of its main advantages is that it generates synthetic examples for the minority class by interpolating between minority-class examples, reducing the risk of overfitting the training data.

Additionally, SMOTE produces synthetic examples similar to -but not identical to- the original minority-class examples, promoting diversity in the training data and improving the classifier’s performance. SMOTE is also straightforward to implement and compatible with many classifiers. It is a desirable option for handling class imbalance in text-classification tasks that often have large and complex datasets.

3.5 Feature Extraction

In our work, we utilized both the unigram approach and the term frequency-inverse document frequency (TF-IDF) technique for feature extraction to highlight the sentiments of the tweets. The unigram approach is a simple, yet effective, method for capturing the essential words in a text corpus. Conversely, TF-IDF assigns weights to words based on their frequency in a document and their rarity in the corpus, which helps distinguish between common and rare words [43]. This technique is advantageous, as it can help identify the essential words in a text corpus and discard noise, resulting in a more meaningful and accurate representation of the data. Therefore, by combining both techniques, we could analyze the sentiment of the tweets in our dataset accurately.

For feature engineering, this research adopts TF-IDF. This approach operates by extracting weighted features from the data and assigning each data term with a few weight values into the model to enhance the performance of ML classifiers [46]. TF-IDF focuses on the most distinctive words, making its integration preferable to overcome the limitation of depending on word counts in SA research. Mathematical functions for TF-IDF are represented in Equations 1 and 2 as follows;

$$\mathrm{tf}(\mathrm{t},\mathrm{d}) = \log(1 + \mathrm{f}_{\mathrm{t,d}}) \quad (1)$$

$$\mathrm{Idf}(\mathrm{t}) = \log\left(\frac{1 + \mathrm{N}}{1 + \mathrm{n}_t}\right) \quad (2)$$

where tf(t,d) represents the count of term t in document d. N represents the total document number and n represents documents containing term t.

Our data is shuffled to make the classification performance more generalizable, reduce the variance and avoid model overfitting. The data is split into 80:20 ratio, where 80% is for training the model and 20% for testing it.

3.6 Classification Methods

In SA, ML classifiers have been utilized in diverse research groups for text classification. These various classifiers have provided different results depending on the applied case study. The ML model could be used for predicting what will happen in the future, learning something from the data or for both uses. First, a good training algorithm is needed to solve the optimization issues and store and process a considerable amount of data. Second, the representation and algorithm solutions for inference must be efficient and effective whenever the model has been properly trained and learned. A learning algorithm's reliability refers to how consistently it produces accurate results over time, even when presented with new data. A reliable algorithm can be trusted to make accurate predictions, even in situations where it has not encountered similar data before. In ML, it is not enough to have a model that can make accurate predictions; it is also important to have a model with a reliable learning algorithm that can continue to provide accurate results as new data becomes available. This is why it is essential to prioritize the reliability of the learning algorithm, even if it means sacrificing some computational resources, such as space and time [47].

All of the ML classifiers that have been covered up to this point produce incredible results across various scenarios, whether they involve SA or other ML context issues. Selecting the best classifier for a given scenario can be a complex and subjective process, as it depends on various factors, such as the size and nature of the dataset, the specific goals of the classification task and the desired number of classes. While a classifier with good performance may seem like the obvious choice, it is not always the case that its performance will remain consistent throughout the training process or when applied to different datasets.

Therefore, it may be necessary to consider multiple classifiers and evaluate their performance on the given dataset through experimentation and statistical analysis. A hypothesis could be formulated based on prior knowledge of the dataset or similar classification tasks, but ultimately, the best classifier must be determined empirically. It is important to note that there is no one-size-fits-all solution for selecting a classifier and the choice should be based on the specific requirements and constraints of the classification task. For that purpose, eight different classifiers will be evaluated side-by-side to check the most accurate one in solving the SA problem in this study. This sub-section will describe the primary classifiers utilized during this work. These classifiers include K-Nearest Neighbor (KNN), Random Forest (RF), Decision Tree (DT), Logistic Regression (LR), Naïve Bayes (NB), XGBoost, AdaBoost and Multi-layer Perception (MLP) classifiers.

3.6.1 KNN

KNN is an essential ML classifier that uses instance-based learning. Text classification uses similarity measurements, which figure out how similar two points are by estimating their distance, proximity or clustering function [48]. In KNN, all training documents are saved and the calculations are postponed up to the classification stage [49]. KNN assigns a class based on the categories of the top neighbors of the labeled samples in the training set for each test document. The closer neighbors with the same category are, the more likely the prediction will be correct [50].

3.6.2 RF

RF is an ensemble classifier that uses bootstrapping and bagging to train several decision trees simultaneously [51]. When using an RF classifier, the final prediction is based on the most commonly observed class of objects and this method is called bagging [52]. A large number of predictors necessitates a lot of planted trees. Individual decision trees can be randomly decorated in various ways; for example, by selecting random features or data sub-sets [53]. To avoid overfitting problems that can occur with individual decision trees because of their tremendous flexibility, RF uses many decision trees on a complex random sub-set of variables to create an effective solution [54].

3.6.3 DT

DT is an ML classification algorithm that works like a hierarchical tree. It utilizes attribute value constraints to split the training data into a few parts and uses different tests to display the tree branch. Each branch slope from the node matches the feature value. A DT works well, as the text-classification model does not have many features, but it is tough to make a classifier when there are a lot of features [55].

3.6.4 LR

LR is considered one of the most prevalent approaches to ML classification [56]. It does this by employing the concept of probability for a single test result by utilizing a logistic function in which the resulting probability might be either 1 or 0 [57]. This methodology has been implemented in various SA research studies [40]. For this reason, it is deemed to be one of the ML classifiers to be evaluated in the classification problem conducted by this research.

3.6.5 NB

NB algorithm is one of the most straightforward examples of a probabilistic classifier [58]. The training documents estimate a class-conditional document distribution, while Bayes' rule is used to get an estimate for test documents. The documents themselves are represented by their words. Furthermore, Naive Bayes may be better than discriminative classifiers for small sample sizes of data, because it has a built-in regularization that makes this method less likely to overfit [59].

3.6.6 XGBoost

Extreme gradient boosting, further called XGBoost, is a powerful ML algorithm that utilizes a gradient-boosting framework to train ensemble models. It works by iteratively adding decision trees to the ensemble, with each new tree correcting the errors of the previous ones, ultimately leading to a more accurate prediction [60]. XGBoost also includes several regularization techniques to prevent overfitting and improve model performance.

3.6.7 AdaBoost

AdaBoost is the first functional boosting classifier suggested by Freund [61]. It combines multiple base classifiers, usually decision trees, to build an accurate classifier. It invokes a weak classifier and provides various training data distributions for each call. The classifier can remove unnecessary features in the training data so that important features are used in the training process. AdaBoost has been utilized in various application studies [62] and has been deemed suitable for the comparative results of this study.

3.6.8 MLP

The MLP is an artificial neural-network architecture, which is probably the most widely used for classification and regression today [63]. MLPs are feed-forward neural networks usually made up of several layers of nodes that only connect in one direction and are usually trained by backpropagation [64].

3.7 Performance Metrics

The performance evaluation measures that are discussed in this research include, Accuracy (Acc.), Precision (Pr.), Recall (Re.), F1 score and Matthews Correlation Coefficient (MCC). These metrics are defined as follows:

$$Accuracy\ (Acc.) = \frac{TP+TN}{TP+FN+FP+TN} \tag{3}$$

$$Precision\ (Pr.) = \frac{TP}{TP+FP} \tag{4}$$

$$Recall\ (Re.) = \frac{TP}{TP+FN} \tag{5}$$

$$F1\ Score = 2x\ \frac{Pr.\,x\ Re.}{Pr.+Re.} \tag{6}$$

$$MCC = \frac{TP\ x\ TN - FP\ x\ FN}{\sqrt{(TP+FP)x\ (TP+FN)x(TN+FP)x(TN+FN)}} \tag{7}$$

TP, FP, TN and FN stand for true positive, false positive, true negative and false negative, respectively.

4. EXPERIMENTAL RESULTS

All classifiers were subjected to two separate experiments: in the first one, the data was imbalanced and in the second one, the imbalanced data was addressed and handled using SMOTE. Table 2 shows the evaluation metrics, including accuracy, precision, recall, F-score and MCC, before applying SMOTE.

Table 2. Results of ML models before applying SMOTE.

Model	Class	Precision	Recall	F-Score	Accuracy	MCC
KNN	0	0.76	0.69	0.72	0.80	0.571
	1	0.83	0.87	0.85		
	Macro avg.	0.79	0.78	0.78		
RF	0	0.95	0.92	0.93	0.95	0.894
	1	0.95	0.97	0.96		
	Macro avg.	0.95	0.94	0.95		
DT	0	0.89	0.92	0.90	0.93	0.847
	1	0.95	0.94	0.94		
	Macro avg.	0.92	0.93	0.92		
LR	0	0.96	0.95	0.95	0.97	0.928
	1	0.97	0.98	0.97		
	Macro avg.	0.97	0.96	0.96		
NB	0	0.85	0.86	0.85	0.89	0.762
	1	0.91	0.91	0.91		
	Macro avg.	0.88	0.88	0.88		
XGBoost	0	0.93	0.83	0.88	0.91	0.816
	1	0.91	0.96	0.93		
	Macro avg.	0.92	0.90	0.91		
AdaBoost	0	0.88	0.54	0.67	0.80	0.572
	1	0.78	0.96	0.96		
	Macro avg.	0.83	0.75	0.76		
MLP	0	0.97	0.97	0.97	0.98	0.956
	1	0.98	0.98	0.98		
	Macro avg.	0.98	0.98	0.98		

In the first experiment, it can be shown that MLP and LR performed superiorly to all of the other ML classifiers in terms of accuracy, with scores of 0.98 and 0.97, respectively. Besides that, RF, DT and XGBoost followed with scores of 0.95, 0.93 and 0.91, correspondingly. Regarding the models on the left side, NB obtained the highest accuracy, which was 0.89, followed by KNN and AdaBoost, with the lowest accuracy of 0.80. In addition to accuracy, the MCC has been recognized in the literature as a comprehensive performance evaluation for binary-classification issues, especially true when using imbalanced and balanced datasets as an evaluation criterion. In this regard, the MCC scored the most for MLP with a value of 0.956, followed by LR with 0.928. However, the score had the lowest for AdaBoost and KNN, with values of 0.572 and 0.571, respectively.

Next, the same classifiers were applied again after balancing the distributed dataset using SMOTE. This experiment was carried out to demonstrate how SMOTE can improve the performance of classifiers after they have been applied to an imbalanced dataset. As they are involved here, earlier employed evaluation measures can also be found in Table 3.

Table 3. Results of ML models after applying SMOTE.

Model	Class	Precision	Recall	F-Score	Accuracy	MCC
KNN	0	0.62	0.99	0.76	0.69	0.423
	1	0.97	0.40	0.57		
	Macro avg.	0.79	0.69	0.66		
RF	0	0.95	0.96	0.96	0.96	0.910
	1	0.96	0.95	0.96		
	Macro avg.	0.96	0.96	0.96		
DT	0	0.95	0.94	0.94	0.94	0.884
	1	0.94	0.95	0.94		
	Macro avg.	0.94	0.94	0.94		
LR	0	0.97	0.97	0.97	0.97	0.944
	1	0.97	0.97	0.97		
	Macro avg.	0.97	0.97	0.97		
NB	0	0.86	0.91	0.89	0.88	0.767
	1	0.91	0.85	0.88		
	Macro avg.	0.88	0.88	0.88		
XGBoost	0	0.95	0.90	0.92	0.93	0.853
	1	0.90	0.95	0.93		
	Macro avg.	0.93	0.93	0.93		
AdaBoost	0	0.92	0.59	0.72	0.77	0.577
	1	0.70	0.95	0.80		
	Macro avg.	0.81	0.77	0.76		
MLP	0	0.99	0.98	0.99	0.99	0.970
	1	0.98	0.99	0.98		
	Macro avg.	0.99	0.99	0.99		

For the following experiment, it can be observed that the best performance was attributed to MLP with 0.99 accuracy, followed by LR, RF, then DT with accuracies of 0.97, 0.96 and 0.94, respectively. Worst accuracy performance was attributed to AdaBoost with 0.77, then KNN with 0.69. As for MCC results, MLP was the highest classifier with 0.97, followed by LR with 0.944, then RF with 0.910. The worst MCC performance was observed at 0.423 in KNN classifier. It is indicated from the results that some classifiers' accuracies have improved after SMOTE was applied to the imbalanced dataset. At the same time, some classifiers' performance has degraded. Still, it is confirmed that MCC across all classifiers has improved, which shows the suitability of SMOTE in performance evaluation after balancing the dataset.

4.1 Comparative Analysis

This sub-section compares the accuracy and MCC values of the results for all ML classifiers before and after using SMOTE. The comparison is illustrated in Figures 5 and 6, respectively.

It is observed that accuracy and MCC are among the most important measures used to evaluate the performance of ML classifiers. Based on the analysis and the analyzed case in this research, it is evident that when the SMOTE technique was applied, the performance of four of the classifiers increased: RF from 0.95 to 0.96, DT from 0.93 to 0.94, XGBoost from 0.91 to 0.93 and MLP from 0.98 to 0.99. With 0.97, only LR kept its accuracy before and after SMOTE. However, the remaining three ML classifiers, Adaboost, KNN and NB, did not demonstrate any gain in accuracy. These results indicate the suitability of the SMOTE technique in terms of accuracy. However, another important measure, MCC introduced in the literature, is more robust and trustworthy than balanced accuracy in F1 score and binary classification analysis [65]. The MCC data shows that most classifiers exhibited an increase after implementing SMOTE, with the most significant improvement reported for MLP 0.97 MCC score, followed by LR 0.944 MCC score. The MCC scores for RF, DT and XGBoost are 0.91, 0.884 and 0.853, respectively. Only the AdaBoost classifier showed a minor gain in the MCC score, bringing it to 0.577. However, the MCC score of the KNN classifier decreased after applying SMOTE, which is also consistent with accuracy.

Figure 7 shows the ROC curve analysis for applied ML models. The results show that MLP, LR and RF have the highest AUC-ROC values, with values of 0.987, 0.988 and 0.987, respectively. These results indicate that these models are more accurate and reliable in predicting the target variable. The KNN model has an AUC-ROC value of 0.863, lower than the other three top-performing models. This suggests that the KNN model may not perform as well in specific scenarios where the other models are better suited. The NB and DT models have lower AUC-ROC values of 0.936 and 0.927, respectively. These values are lower than for the top-performing models, indicating that these models may not be as accurate in predicting the target variable as the others. Lastly, Adaboost and XGB models have AUC-ROC values of 0.865 and 0.965, respectively. While the AUC-ROC value for Adaboost is lower than most other models, the XGB model has a relatively high AUC-ROC value, suggesting that it may be a good alternative to the top-performing models.

Overall, the AUC-ROC values give a good indication of the relative performance of each classification model and can be used to guide our choice of the best model for our specific classification problem.

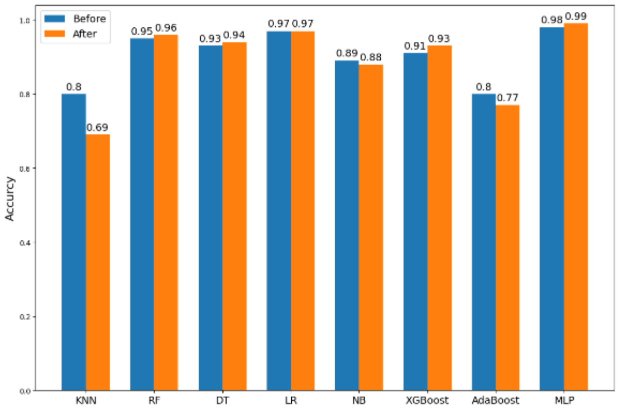

Figure 5. Comparative analysis of accuracy before and after applying SMOTE.

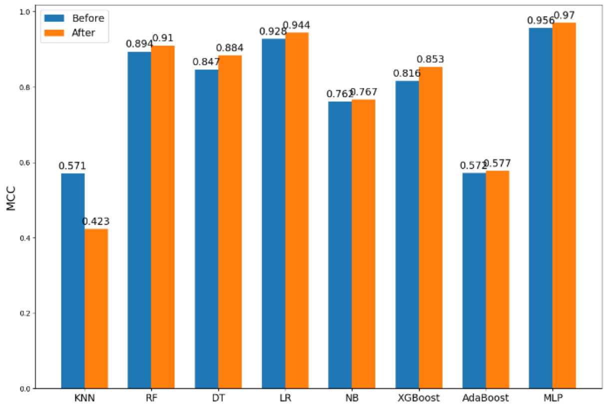

Figure 6. Comparative analysis of MCC before and after applying SMOTE.

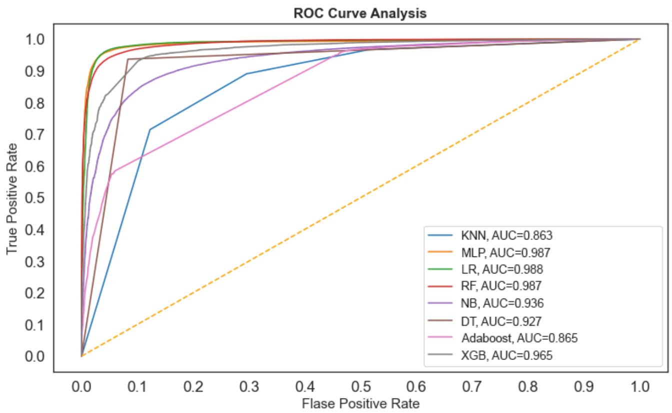

Figure 7. ROC curve analysis for applied ML models.

4.2 Result Discussion

In the classification of Russian aggression in Ukraine-related discussion on Twitter, it is evident that most basic ML classifiers improved their performance, which was confirmed by measuring the MCC score as identified in the literature to be one of the best approaches for classification problems, particularly when data is balanced utilizing techniques, such as SMOTE. The only classifier not enhanced by the used approach was the KNN classifier, validated by the MCC score and the accuracy result. Even so, the KNN algorithm performed far higher when the data was imbalanced than when the data was balanced. This demonstrates that despite the promise of data-balancing methodologies, their application in producing a balanced dataset could not always be applicable across all ML classifiers. As a result, it is worthwhile to investigate the possibility of determining the performance of these various classifiers by employing additional data-balancing methods to evaluate and compare their performance.

In this study, the numerical results obtained from the ML models should be discussed to provide insights into the performance of the models. The results presented in Table 1 and Table 2 show the performance of the models before and after applying SMOTE, respectively. Before applying SMOTE, the RF and

MLP models had the highest precision, recall, F-score, accuracy and MCC values, indicating that they performed the best among the models. After applying SMOTE, the RF and MLP models still had the highest values for most of these metrics, indicating that they continued to perform well even after the dataset was balanced using SMOTE.

In contrast, the KNN and AdaBoost models had lower performances before applying SMOTE, with lower precision, recall, F-score, accuracy and MCC values. After using SMOTE, these models showed some performance improvement, but did not perform as well as the RF and MLP models. The DT and NB models had moderate performances before and after applying SMOTE, with relatively consistent values for most metrics. These results can be discussed regarding the strengths and weaknesses of the different models and how well they handled the imbalanced dataset. Additionally, the implications of these results for the problem being addressed in the study can be discussed, including any recommendations for selecting a model or improving the performances of the models.

Table 4. Comparison with existing Twitter sentiment classification methods.

Reference	**Twitter Sentiment Classifier**	**Accuracy (%)**
[66] 2018	RNN-Capsule	91.6%
[67] 2019	Hybrid CNN-LSTM model	91%
[68] 2021	ConvBiLSTM model	91.3%
Our best model	MLP	98%

Table 4 provides a comparison of the performance of the proposed MLP model with those of three existing Twitter sentiment classification methods. The table reports the accuracy of each method as a percentage. It can be observed that the proposed MLP model outperforms the existing methods, achieving an accuracy of 98%. In contrast, the existing methods report 91% to 91.6% accuracy. The comparison of the proposed MLP model with the current methods demonstrates the effectiveness of the proposed approach in classifying Twitter sentiments. The MLP model outperforms the existing methods by a significant margin, indicating that the proposed approach can improve the accuracy of Twitter sentiment classification. However, it is essential to note that the comparison is limited to the reported accuracy metric. Other evaluation metrics, such as precision, recall, F1 score and ROC-AUC, should also be considered to evaluate the proposed method comprehensively. Overall, the results in Table 4 suggest that the proposed MLP model is a promising approach for Twitter sentiment classification and can provide improved accuracy compared to existing methods.

5. CONCLUSION

In this study, sentiments about war during the Russian aggression in Ukraine have been analyzed. This study achieved two goals: the uniqueness of the collected data and ML to categorize the tweets' sentiments. The first goal was to collect the dataset by searching for the most popular hashtags about aggression. The second goal was to place the collected tweets into categories using several well-known ML models. The most basic ML classifiers improved their performance, confirmed by evaluating the MCC score, which is known in the literature as one of the best ways to solve classification problems, especially when data is balanced using techniques like SMOTE. Also, it was demonstrated that data-balancing techniques would not guarantee that all classes could perform better. Nevertheless, the data-balancing approach must be tested and compared using different ML classifiers and SA evaluation datasets.

The prediction of sentiment analysis on Russian aggression in Ukraine using ML models has significant implications for the academia. Firstly, it can enhance our understanding of the social and emotional aspects of aggression, particularly the sentiments of those affected by it. By predicting sentiment trends over time, researchers can identify patterns in public opinion and gain insights into the underlying causes and factors that contribute to positive or negative sentiments.

Secondly, using ML models for sentiment analysis can provide a more accurate and efficient analysis of large volumes of data. Traditional manual sentiment-analysis methods can be time-consuming and subjective, leading to potential biases and errors. ML models, on the other hand, can analyze large datasets in real time, providing quick and accurate results.

Furthermore, using ML models for sentiment analysis can provide valuable insights for policymakers and decision-makers. Policymakers can develop more effective conflict-resolution and peace-building strategies by identifying potential sources of tension and aggression escalation. Moreover, ML models can help determine the effectiveness of propaganda and messaging campaigns used by the parties involved, which can aid in designing more effective messaging strategies.

On the other hand, there are several limitations to consider in the study. Firstly, shortened links and multimedia content were not considered, leading to underestimating Russian propaganda and other sources. Secondly, the study relied on a distant-supervision approach rather than manual verification, which could introduce errors and biases. Additionally, the method used to assess the amount of removed content was imperfect and did not allow for the exact reasons for removal. The study did not account for the activity of automated accounts that could spread misinformation.

In summary, the prediction of sentiment analysis on Russian aggression in Ukraine using ML models can advance our understanding of the conflict's social and emotional aspects, provide an accurate and efficient analysis of large volumes of data and aid policymakers in developing more effective conflict-resolution strategies. As such, it is an important area of research for the academia.

In the future, the current research can be expanded by incorporating deep-learning classifiers, exploring various feature settings, experimenting with different data-balancing techniques and conducting more predictive analysis research on both the SA dataset presented here and other benchmarking datasets from the research literature. These future works have the potential to enhance our technical understanding of ML and its configurations and parameters and provide us with deeper insights into the performance of ML models in sentiment-analysis tasks. By further exploring these avenues, we can better understand the strengths and limitations of different ML algorithms and techniques and identify more effective ways to optimize their performance. Overall, these future works will help advance the field of ML and its application in sentiment analysis and open up new avenues for future research.

Compliance with Ethical Standards

We certify that our work is original and does not plagiarize the work of others.

Competing Interests

We certify that there is no actual or potential conflict of interest in relation to this article.

Research Data Policy and Data Availability Statements

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

We would like to thank Dr. Ahmed M. Fakhrudeen from the University of Kirkuk for his invaluable contribution to this manuscript. His efforts in proofreading and providing feedback have significantly improved its quality. We appreciate his dedication to the field and extend our heartfelt thanks for his valuable support in bringing this manuscript to fruition.

References

[1] A. H. Alamoodi, M. R. Baker, O. S. Albahri, B. B. Zaidan and A. A. Zaidan, "Public Sentiment Analysis and Topic Modeling Regarding COVID-19's Three Waves of Total Lockdown: A Case Study on Movement Control Order in Malaysia," KSII Trans. Internet Inf. Syst., vol. 16, no. 7, pp. 2169–2190, DOI: 10.3837/tiis.2022.07.003, 2022.

[2] N. Afroz, M. Boral, V. Sharma and M. Gupta, "Sentiment Analysis of COVID-19 Nationwide Lockdown Effect in India," Proc. of the Int. Conf. on Artificial Intelligence and Smart Systems (ICAIS 2021), pp. 561–567, DOI: 10.1109/ICAIS50930.2021.9396038, 2021.

[3] S. Hajrahnur, M. Nasrun, C. Setianingsih and M. A. Murti, "Classification of Posts on Twitter Traffic Jam in the City of Jakarta Using Algorithm C4.5," Proc. of the 2018 Int. Conf. on Signals and Systems (ICSigSys 2018), pp. 294–300, DOI: 10.1109/ICSIGSYS.2018.8372776, 2018.

[4] P. Kostakos, M. Nykanen, M. Martinviita, A. Pandya and M. Oussalah, "Meta-terrorism: Identifying Linguistic Patterns in Public Discourse After an Attack," Proc. of the 2018 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining (ASONAM 2018), pp. 1079–1083, DOI:

10.1109/ASONAM.2018.8508647, 2018.

[5] G. M. Demirci, S. R. Keskin and G. Dogan, "Sentiment Analysis in Turkish with Deep Learning," Proc. of the 2019 IEEE Int. Conf. on Big Data (Big Data 2019), pp. 2215–2221, DOI: 10.1109/BigData47090.2019.9006066, 2019.

[6] J. P. Singh, Y. K. Dwivedi, N. P. Rana, A. Kumar and K. K. Kapoor, "Event Classification and Location Prediction from Tweets during Disasters," Annals of Operations Research, vol. 283, no. 1–2, pp. 737–757, DOI: 10.1007/s10479-017-2522-3, Dec. 2019.

[7] N. H. Khun, T. T. Zin, M. Yokota and H. A. Thant, "Emotion Analysis of Twitter Users on Natural Disasters," Proc. of the 2019 IEEE 8th Global Conf, on Consumer Electronics (GCCE 2019), pp. 342–343, DOI: 10.1109/GCCE46687.2019.9015234, 2019.

[8] U. H. H. Zaki, R. Ibrahim, S. A. Halim, K. A. M. Khaidzir and T. Yokoi, "Sentiflood: Process Model for Flood Disaster Sentiment Analysis," Proc. of the 2017 IEEE Conf. on Big Data and Analytics (ICBDA 2017), vol. 2018-Janua., pp. 37–42, DOI: 10.1109/ICBDAA.2017.8284104, 2018.

[9] S. K. Akpatsa et al., "Sentiment Analysis and Topic Modeling of Twitter Data: A Text Mining Approach to the US-Afghan War Crisis," SSRN Electronic J., DOI: 10.2139/ssrn.4064560, 2022.

[10] E. Lee, F. Rustam, I. Ashraf, P. B. Washington, M. Narra and R. Shafique, "Inquest of Current Situation in Afghanistan under Taliban Rule Using Sentiment Analysis and Volume Analysis," IEEE Access, vol. 10, pp. 10333–10348, DOI: 10.1109/ACCESS.2022.3144659, 2022.

[11] M. Mahiuddin, "Real Time Sentiment Analysis and Opinion Mining on Refugee Crisis," Proc. of the 2019 5th Int. Conf. on Advances in Electrical Engineering (ICAEE 2019), pp. 699–705, DOI: 10.1109/ICAEE48663.2019.8975462, 2019.

[12] A. Alamoodi et al., "Sentiment Analysis and Its Applications in Fighting COVID-19 and Infectious Diseases: A Systematic Review," Expert Systems with Applications, vol. 167, p. 114155, 2020.

[13] G. Assembly, "Aggression against Ukraine: Resolution / Adopted by the General Assembly," United Nations, [Online], Available: https://digitallibrary.un.org/record/3959039?ln=en, 2022.

[14] M. M. Metzger, R. Bonneau, J. Nagler and J. A. Tucker, "Tweeting Identity? Ukrainian, Russian and# Euromaidan," J. of Comparative Economics, vol. 4, no. 1, pp. 16–40, 2016.

[15] J. Driscoll and Z. C. Steinert-Threlkeld, "Social Media and Russian Territorial Irredentism: Some Facts and a Conjecture," Post-Soviet Aff., vol. 36, no. 2, pp. 101–121, Mar. 2020.

[16] R. A. Bryant, P. P. Schnurr and D. Pedlar, "Addressing the Mental Health Needs of Civilian Combatants in Ukraine," The Lancet Psychiatry, vol. 9, no. 5, pp. 346–347, 2022.

[17] E. Elmurngi and A. Gherbi, "Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques," Proc. of the 6th Int. Conf. Data Analytics Detection (DATA Anal. 2017), no. c, pp. 65–72, 2017.

[18] W. F. Al-Sarraj and H. M. Lubbad, "Bias Detection of Palestinian/Israeli Conflict in Western Media: A Sentiment Analysis Experimental Study," Proc. of the 2018 Int. Conf. on Promising Electronic Technologies (ICPET 2018), pp. 98–103, DOI: 10.1109/ICPET.2018.00024, 2018.

[19] N. Öztürk and S. Ayvaz, "Sentiment Analysis on Twitter: A Text Mining Approach to the Syrian Refugee Crisis," Telematics and Informatics, vol. 35, no. 1, pp. 136–147, DOI: 10.1016/j.tele.2017.10.006, 2018.

[20] S. Mansour, "Social Media Analysis of Users' Responses to Terrorism Using Sentiment Analysis and Text Mining," Procedia Computer Science, vol. 140, pp. 95–103, DOI: 10.1016/j.procs.2018.10.297, 2018.

[21] G. A. Ruz, P. A. Henríquez and A. Mascareño, "Sentiment Analysis of Twitter Data During Critical Events through Bayesian Networks Classifiers," Future Generation Computer Systems, vol. 106, pp. 92–104, DOI: 10.1016/j.future.2020.01.005, 2020.

[22] F. Yao and Y. Wang, "Domain-specific Sentiment Analysis for Tweets during Hurricanes (DSSA-H): A Domain-adversarial Neural-network-based Approach," Computers, Environment and Urban Systems, vol. 83, DOI: 10.1016/j.compenvurbsys.2020.101522, 2020.

[23] A. Squicciarini, A. Tapia and S. Stehle, "Sentiment Analysis during Hurricane Sandy in Emergency Response," Int. J. of Disaster Risk Reduct., vol. 21, pp. 213–222, DOI: 10.1016/j.ijdrr.2016.12.011, 2017.

[24] S. H. W. Ilyas, Z. T. Soomro, A. Anwar, H. Shahzad and U. Yaqub, "Analyzing Brexit's Impact Using Sentiment Analysis and Topic Modeling on Twitter Discussion, " Proc. of the ACM Int. Conf., pp. 1–6, DOI: 10.1145/3396956.3396973, Jun. 2020.

[25] A. Field, C. Y. Park, A. Theophilo, J. Watson-Daniels and Y. Tsvetkov, "An Analysis of Emotions and the Prominence of Positivity in #BlackLivesMatter Tweets," Proc. of the National Academy of Sciences of the United States of America, vol. 119, no. 35, p. e2205767119, DOI: 10.1073/pnas.2205767119, 2022.

[26] D. Won, Z. C. Steinert-Threlkeld and J. Joo, "Protest Activity Detection and Perceived Violence Estimation from Social Media Images," Proc. of the 2017 ACM Multimedia Conf.e (MM 2017), pp. 786–794, DOI: 10.1145/3123266.3123282, Oct. 2017.

[27] Z. Steinert-Threlkeld and J. Joo, "MMCHIVED: Multimodal Chile and Venezuela Protest Event Data," Proc. of the 16th Int. AAAI Conf. on Web and Social Media, vol. 16, pp. 1332–1341, DOI: 10.1609/icwsm.v16i1.19385, 2022.

[28] F. Rustam, M. Khalid, W. Aslam, V. Rupapara, A. Mehmood and G. S. Choi, "A Performance Comparison

of Supervised Machine Learning Models for Covid-19 Tweets Sentiment Analysis," PLoS One, vol. 16, no. 2, DOI: 10.1371/journal.pone.0245909, Feb. 2021.

[29] Imamah and F. H. Rachman, "Twitter Sentiment Analysis of Covid-19 Using Term Weighting TF-IDF and Logistic Regresion," Proc. of the 6th Information Technology Int. Seminar (ITIS 2020), pp. 238–242, DOI: 10.1109/ITIS50118.2020.9320958, 2020.

[30] P. Sharma and A. K. Sharma, "Experimental Investigation of Automated System for Twitter Sentiment Analysis to Predict the Public Emotions Using Machine Learning Algorithms," Materials Today Proc., DOI: 10.1016/j.matpr.2020.09.351, 2020.

[31] M. Caprolu, A. Sadighian and R. Di Pietro, "Characterizing the 2022 Russo-Ukrainian Conflict through the Lenses of Aspect-based Sentiment Analysis: Dataset, Methodology and Preliminary Findings," arXiv Prepr, arXiv2208.04903, [Online], Available: http://arxiv.org/abs/2208.04903, Aug. 2022.

[32] H. W. A. Hanley, D. Kumar and Z. Durumeric, "A Special Operation': A Quantitative Approach to Dissecting and Comparing Different Media Ecosystems' Coverage of the Russo-Ukrainian War," arXiv Prepr, arXiv 2210.03016, [Online], Available: https://doi.org/10.48550/arXiv.2210.03016, Oct. 2022.

[33] A. Guerra and O. Karakuş, "Sentiment Analysis for Measuring Hope and Fear from Reddit Posts during the 2022 Russo-Ukrainian Conflict," arXiv Prepr, arXiv2301.08347, [Online], Available: http://arxiv.org/abs/2301.08347, Jan. 2023.

[34] F. Pierri, L. Luceri, N. Jindal and E. Ferrara, "Propaganda and Misinformation on Facebook and Twitter during the Russian Invasion of Ukraine," arXiv Prepr, arXiv2212.00419, Accessed: Apr. 04, 2023. [Online], Available: http://arxiv.org/abs/2212.00419, Dec. 2022.

[35] N. S. Agarwal, N. S. Punn and S. K. Sonbhadra, "Exploring Public Opinion Dynamics on the Verge of World War III Using Russia-Ukraine War-Tweets Dataset," KDD-UC, Washington, DC, USA, [Online], Available: https://www.kdd.org/kdd2022/papers/27_Navya Sonal Agarwal.pdf, 2022.

[36] P. Vyas, M. Reisslein, B. P. Rimal, G. Vyas, G. P. Basyal and P. Muzumdar, "Automated Classification of Societal Sentiments on Twitter with Machine Learning," IEEE Transactions on Technology and Socity, vol. 3, no. 2, pp. 100–110, DOI: 10.1109/tts.2021.3108963, 2021.

[37] R. Ibar-Alonso, R. Quiroga-García and M. Arenas-Parra, "Opinion Mining of Green Energy Sentiment: A Russia-Ukraine Conflict Analysis," Mathematics, vol. 10, no. 14, DOI: 10.3390/math10142532, 2022.

[38] B. Chen et al., "Public Opinion Dynamics in Cyberspace on Russia-Ukraine War: A Case Analysis with Chinese Weibo," IEEE Transactions on Computational Social Systems, vol. 9, no. 3, pp. 948–958, 2022.

[39] M. B. Garcia and A. Cunanan-Yabut, "Public Sentiment and Emotion Analyses of Twitter Data on the 2022 Russian Invasion of Ukraine," Proc. of the 2022 9th Int. Conf. on Information Technology, Computer and Electrical Engineering (ICITACEE 2022), pp. 242–247, DOI: 10.1109/ ICITACEE55701.2022.99241 36, 2022.

[40] B. Džubur, Ž. Trojer and U. Zrimšek, "Semantic Analysis of Russo-Ukrainian War Tweet Networks," SCORES: Ljubljana, [Online], Available: http://www.scores.si/assets/papers/6258.pdf, 2022.

[41] Z. C. Steinert-Threlkeld, Twitter As Data, DOI: 10.1017/9781108529327, Cambridge Uni. Press, 2018.

[42] Mendeley Data, "Russian Aggression in Ukraine Related Tweets - Mendeley Data," DOI:10.17632/77xdt9 25zp.1, 2023.

[43] M. R. Baker and M. A. Akcayol, "A Novel Web Ranking Algorithm Based on Pages Multi-attribute," Int. J. of Information Technology, vol. 14, no. 2, pp. 739–749, DOI: 10.1007/s41870-021-00833-5, 2022.

[44] A. Krouska, C. Troussas and M. Virvou, "The Effect of Preprocessing Techniques on Twitter Sentiment Analysis," Proc. of the 7th Int. Conf. on Information, Intelligence, Systems and Applications (IISA 2016), pp. 1–5. DOI: 10.1109/IISA.2016.7785373, Dec. 2016.

[45] M. A. Abid, S. Ullah, M. A. Siddique, M. F. Mushtaq, W. Aljedaani and F. Rustam, "Spam SMS Filtering Based on Text Features and Supervised Machine Learning Techniques," Multimedia Tools and Applications, vol. 81, pp. 39853–39871, DOI: 10.1007/s11042-022-12991-0, 2022.

[46] K. Chen, Z. Zhang, J. Long and H. Zhang, "Turning from TF-IDF to TF-IGM for Term Weighting in Text Classification," Expert Systems with Applications, vol. 66, DOI: 10.1016/j.eswa.2016.09.009, 2016.

[47] E. Alpaydin, Introduction to Machine Learning, 4th Edn., MIT Press, DOI: 10.1007/978-3-030-74640-7_4, 2020.

[48] V. K. Vijayan, K. R. Bindu and L. Parameswaran, "A Comprehensive Study of Text Classification Algorithms," Proc. of the 2017 Int. Conf. on Advances in Computing, Communications and Informatics (ICACCI 2017), vol. 2017-Jan., pp. 1109–1113, DOI: 10.1109/ICACCI.2017.8125990, 2017.

[49] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, DOI: 10.1145/505282.505283, 2002.

[50] Y. Yang and X. Liu, "A Re-examination of Text Categorization Methods," Proc. of the 22nd Annual Int. ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 1999), pp. 42–49, DOI: 10.1145/312624.312647, Aug. 1999.

[51] N. Jalal, A. Mehmood, G. S. Choi and I. Ashraf, "A Novel Improved Random Forest for Text Classification Using Feature Ranking and Optimal Number of Trees," J. King Saud Univ. - Comput. Inf. Sci., DOI: 10.1016/j.jksuci.2022.03.012, 2022.

[52] L. Breiman, "Bagging Predictors," Machine Learning, vol. 24, no. 2, pp. 123–140, 1996.
[53] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[54] P. Domingos, "A Few Useful Things to Know about Machine Learning," Communications of the ACM, vol. 55, no. 10, pp. 78–87, DOI: 10.1145/2347736.2347755, Oct. 2012.
[55] B. Agarwal and N. Mittal, "Text Classification Using Machine Learning Methods: A Survey," Advances in Intelligent Systems and Comp., vol. 236, pp. 701–709, DOI: 10.1007/978-81-322-1602-5_75, 2014.
[56] A. Subasi, Practical Machine Learning for Data Analysis Using Python, Elsevier, DOI: 10.1016/B978-0-12-821379-7.00008-4, 2020.
[57] H. Belyadi and A. Haghighat, Machine Learning Guide for Oil and Gas Using Python, Elsevier, DOI: 10.1016/c2019-0-03617-5, 2021.
[58] Y. Yang, "An Evaluation of Statistical Approaches to Text Categorization," Information Retrieval, vol. 1, no. 1–2, pp. 69–90, DOI: 10.1023/a:1009982220290, 1999.
[59] S. Wang and C. D. Manning, "Baselines and Bigrams: Simple, Good Sentiment and Topic Classification," Proc. of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012), vol. 2, pp. 90–94, 2012.
[60] R. Can, S. Kocaman and C. Gokceoglu, "A Comprehensive Assessment of XGBoost Algorithm for Landslide Susceptibility Mapping in the Upper Basin of Ataturk Dam, Turkey," Applied Sciences, vol. 11, no. 11, p. 4993, DOI: 10.3390/app11114993, 2021.
[61] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm," Proc. of the 13th Int. Conf. Machine Learning, pp. 148–156, DOI: 10.1.1.133.1040, 1996.
[62] W. Wang and D. Sun, "The Improved AdaBoost Algorithms for Imbalanced Data Classification," Information Sciences, vol. 563, pp. 358–374, DOI: 10.1016/j.ins.2021.03.042, Jul. 2021.
[63] A. Diera et al., "Bag-of-Words *vs.* Sequence *vs.* Graph *vs.* Hierarchy for Single- and Multi-label Text Classification," Proc. of the 60th Annual Meeting of the Association for Computational Linguistics, vol. 1: Long Papers, pp. 4038 - 4051, DOI: 10.48550/arXiv.2204.03954, 2022.
[64] A. Pinkus, "Approximation Theory of the MLP Model in Neural Networks," Acta Numerica, vol. 8, pp. 143–195, DOI: 10.1017/S0962492900002919, 1999.
[65] D. Chicco and G. Jurman, "The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation," BMC Genomics, vol. 21, no. 1, pp. 1–13, DOI: 10.1186/s12864-019-6413-7, Jan. 2020.
[66] Y. Wang, A. Sun, J. Han, Y. Liu and X. Zhu, "Sentiment Analysis by Capsules," Proc. of the World Wide Web Conference (WWW 2018), vol. 10, pp. 1165–1174, DOI: 10.1145/3178876.3186015, Apr. 2018.
[67] A. U. Rehman, A. K. Malik, B. Raza and W. Ali, "A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis," Multimedia Tools and Applications, vol. 78, no. 18, pp. 26597–26613, DOI: 10.1007/s11042-019-07788-7, Sep. 2019.
[68] S. Tam, R. Ben Said and Ö. Tanriöver, "A ConvBiLSTM Deep Learning Model-based Approach for Twitter Sentiment Classification," IEEE Access, vol. 9, pp. 41283–41293, DOI: 10.1109/ACCESS.2021.3064830, 2021.

ملخص البحث:

تبحث الدّراسة في الكيفية التي عبّر بها النّاس عن مشاعرهم على "تويتر" خلال حرْب روسيا على أوكرانيا. وحققت الدّراسة غرضين؛ فقد جمعت معلوماتٍ فريدة، كما استخدمت تعلُّم الآلة لتصنيف التّغريدات بناءً على أثرها على أحاسيس النّاس. وقد جَهِدت الدّراسة لإيجاد أكثر "الهاشتاغات" علاقةً بالحرب الرّوسية-الأوكرانية من أجل تحديد مجموعة البيانات الخاصّة بالدّراسة. كذلك عملَتْ على استخدام عددٍ من نماذج تعلُّم الآلة بغية تنظيم التّغريدات في مجموعتين.

وقد بيّنت النّتائج التّجريبية أنّ غالبية مُصنِّفات تعلُّم الآلة المستخدمة كانت ذات دقّةٍ أعلى عند استخدام مجموعات بيانات متوازنة. إلّا أنّه اتّضح أنّ استراتيجيات موازنة البيانات لم تكُنْ كلّها بالنّجاعة ذاتها. لذا كان لا بدّ من تسْليط الضّوء على مقارنة تلك الاستراتيجيات المستخدمة لموازنة البيانات في تحليل المشاعِر باستخدام تعلُّم الآلة، وذلك عبر استخدام مصنِّفات أكثر واختبارها في مدى واسعٍ من المهامّ.