Sentiment Analysis of Twitter Posts on Global Conflicts

Ujwal Sasikumar
School of Workforce Development
Conestoga College ITAL
Kitchener ON, Canada
ujusasikumar@gmail.com

ANK Zaman
Dept. of Physics and Computer Science
Wilfrid Laurier University
Waterloo ON, Canada
azaman@wlu.ca

Abdul-Rahman Mawlood-Yunis
Dept. of Physics and Computer Science
Wilfrid Laurier University
Waterloo ON, Canada
amawloodyunis@wlu.ca

Prosenjit Chatterjee
Dept. of Computer Science and Cyber Security
Southern Utah University
Cedar City, UT, USA
prosenjitchatterjee@suu.edu

Abstract—Sentiment analysis of social media data is an emerging field with vast applications in various domains. In this study, we developed a sentiment analysis model to analyze social media sentiment, especially tweets, during global conflicting scenarios. To establish our research experiment, we identified a recent global dispute incident on Twitter and collected around 31,000 filtered Tweets for several months to analyze human sentiment worldwide. In this research, the training models comprised of Naïve Bayes Algorithm (NBA) and Neural Networks (NNs). The trained model achieved 80% and 100% accuracy in predicting the sentiment of the tweets using NBA and NNs models, respectively. A confusion matrix was created to evaluate the model's performance, indicating that the model correctly classified many positive and negative tweets. This study contributes to develop sentiment analysis models for analyzing social media data related to various events and topics. The findings can be helpful for researchers and practitioners interested in understanding public sentiment about the country-wide conflict on social media. In this research, Twitter is referred to its new name X.

Index Terms—Sentiment Analysis, Twitter, X, Machine Learning, Naive Bayes, Neural Network, Affin.

I. INTRODUCTION

The complex geopolitical war between countries, which has enormous worldwide repercussions, has drawn much interest on social media platforms, particularly Twitter. This study thoroughly investigates the opinions posted on Twitter about the current warfare in the area. This work examines a large corpus of tweets using cutting-edge NLP and machine-learning (ML) approaches to find the prevailing sentiment trends, viewpoints, and feelings around any complex conflict. Modern sentiment analysis methods and procedures are used in our study to classify Twitter corpus into positive, negative, and neutral sentiment categories. From this analysis, we want to learn how people, groups, and the global community view and respond to international conflict. This work also investigates temporal trends and regional differences in sentiment expression, offering information on how public opinion transformed throughout the conflict.

A valuable resource for policymakers, researchers, and the general public looking to understand the tones [1] of public opinion and sentiment surrounding this ongoing global issue is provided by this study, which contributes to a deeper understanding of the social and emotional aspects of the international War between two countries. Social media has become an increasingly important source of information during major events and crises, including wars and conflicts [2]. Twitter has been extensively used among social media platforms to share news, opinions, and emotions about ongoing events [3]. The recent global conflict in the year 2022-23 between multiple countries is one such event that has garnered significant attention on social media, with users from around the world expressing their views and sentiments [4], [5]. This study aims to build a sentiment analysis model to determine the sentiment of tweets during the said war posted by independent country (not involved in any war) users. Sentiment analysis is a subfield of natural language processing (NLP) that involves using computational techniques to extract and quantify subjective information from text [4]. NLP is a rapidly growing field of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human language [6]. By applying NLP techniques, we aim to develop a model that can accurately classify tweets as positive, negative, or neutral based on the sentiment expressed in the text. To achieve this goal, the NBA and NNs are popular ML techniques extensively used for text classification [7]. The algorithms learn the probability of each object's features and classify them into several predetermined categories. Supervised learning is the approach to train our model, using a set of annotated tweets to identify the relationship between the tweet's features and its corresponding sentiment. The main objective of this study is to evaluate the accuracy of the developed model in classifying

the sentiment of tweets related to the international war posted by independent country's users. The results of this study could have significant implications for understanding the impact of social media on public opinion and sentiment during times of global conflict. Researchers have used a variety of approaches and technology in recent years to do sentiment analysis on Twitter data related to the war-time conflict scenario. This literature review delves into the most important studies and conclusions. Serpil Aslan [8] examines sentiment analysis in the context of the Ukraine situation. It analyses public opinion using Twitter data and exploring users' emotional responses to any war, illuminating how emotions change over time. The rest of the paper is organized as follows: Section II describes the dataset used in this research, Section III represents the technical details of the implementation, Section IV represents results, and Section V describes the concluding remarks of this paper.

II. DATASET

The following sections present the data collection and preprocessing steps to prepare the dataset for evaluation.

A. Data Collection

The data used for building the sentiment analysis model was collected from Canadian independent country's users' tweets during the international wartime in the year 2022. The tweets were obtained using the Twitter API [9] and stored in a CSV file. Additionally, we utilized Snscrape, a Python-based web scraper, to gather information from Twitter for instances that did not require API use. We leveraged the NBA and NNs, popular machine learning algorithms, to analyze the collected tweets. Specifically, we utilized the Scikit-learn (sklearn) library [10] [11], which is a free open-source machine-learning library for the Python programming language. Scikit-learn offers a variety of features such as classification, regression, and clustering, and we employed it for classification purposes. Each tweet in our dataset[1] [12] was labeled as positive, neutral or negative based on the sentiment expressed in the text. To accomplish this, we used the Afinn library [13], which assigns a sentiment score ranging from (-1) to (+1) to each tweet, with (-1) representing the most negative sentiment and (+1) representing the most positive. In addition, we filtered out tweets that did not contain relevant information to ensure the quality and validity of our dataset.

B. Data Preprocessing

Data preprocessing is critical to any ML and/or NLP activity. It involves transforming raw data into a format that ML algorithms can easily analyze and understand. Our study used data preprocessing on the twitter dataset to prepare the data for sentiment analysis. The first step in data preprocessing was removing irrelevant data, such as numbers and punctuation marks. To achieve the preprocessing goal, regular expression was used in identifying unwanted text for

elimination [14]. In NLP research, frequently used words are called stop words [15] and are removed as they do not have significance in the context of sentiment analysis. This step was performed using the Natural Language Toolkit (NLTK) library, which provides access to predefined stop words for the English text [16]. In our second step, the text was converted to lowercase to ensure consistency in the dataset and avoid any inconsistencies arising from uppercase and lowercase characters in the text data [17]. Stemming [18] was also utilized to reduce each word to its base or root form to further reduce the dimensionality of the dataset, making it easier to analyze and understand. Parts-of-speech (POS) tagging and chunking were also employed to identify the grammatical structure of each sentence. This technique involved assigning POS tags, such as nouns, verbs, adjectives, or adverbs, to each word in the text data. The words were then processed together into adjacent words in a sentence into phrases based on their POS tags. The data processing phase was carried out using several libraries, including Pandas for reading the CSV file, NumPy for numerical computations, Seaborn and Matplotlib for data visualization, and Scikit-learn for ML algorithms [10]. A popularly available cloud server platform was utilized to conduct the experiment. The preprocessing techniques used in this study helped clean and transform the tweets dataset into a format easily analyzed and understood by ML algorithms. This step was critical in ensuring the accuracy and effectiveness of the sentiment analysis.

C. Data Visualization

After performing the data cleaning, we visualized the cleaned data for further processing. Figure 1 displays the sample length of tweets of the dataset.
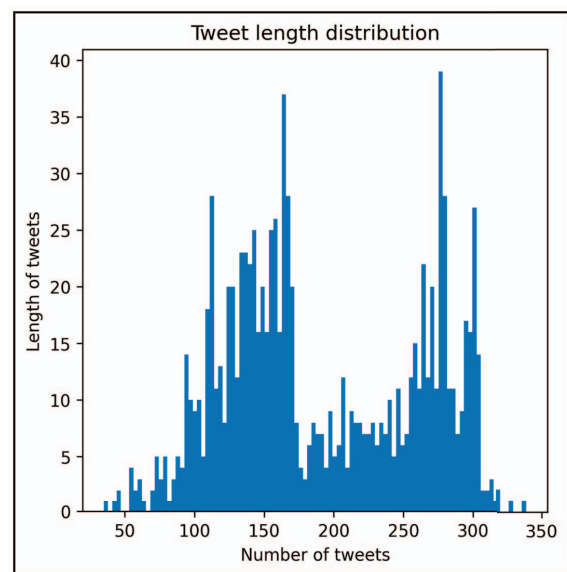

Fig. 1. Tweet length distribution

Figure 2 represents the neutral tweet words from the dataset using a word cloud.

Fig. 2. Word cloud for neutral tokens

Figure 4 represents the positive tweet words from the dataset using a word cloud.

Fig. 3. Word cloud for positive tokens

Figure 4 represents the negative tweet words from the dataset using a word cloud.
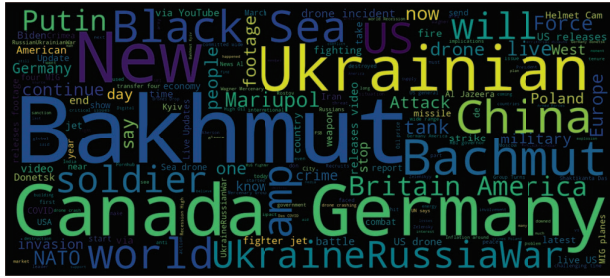

Fig. 4. Word cloud for negative tokens

III. IMPLEMENTATION

The subsequent sections illustrate the process of building and training models to evaluate a sentiment dataset.

A. Naïve Bayes Classifier Model

We used the Naive Bayes (NB) [19] classifier, a popular machine learning algorithm as a text classifier, to train our sentiment analysis model. The Naive Bayes classifier selects the class with the highest probability by first calculating the probability of each class based on the input features [16]. We

split the dataset into training and testing sets, utilizing 80% for training and 20% for testing purposes.

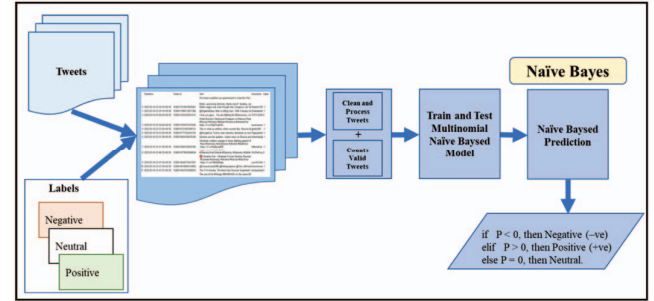Figure 5 shows the NB classifier model, and its processing steps.

Fig. 5. Naïve Bayes Classifier Model

For the purpose of ML data preprocessing we removed punctuations and stopwords from the tweet text. We used the CountVectorizer [20] to convert the text data into a bag-of-words format, which represents the occurrence of each word in each tweet. The resulting matrix was split into features and labels. The Naive Bayes classifier was then trained using the training data, and its accuracy was evaluated using the testing data. To evaluate the performance of the model, we generated a confusion matrix and classification report. The confusion matrix shows the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), while the classification report shows metrics such as precision, recall, and f1 score for each class.

Precision is the ratio of true positives (TP) to the sum of true positives and false positives (FP). Recall: This is the ratio of true positives (TP) to the sum of true positives and false negatives (FN). f1-Score is the harmonic mean of precision and recall and is used as a single metric that balances the trade-off between precision and recall. The F1score ranges from 0 to 1, with 1 being the best possible score.

The formal definitions of the mentioned terms are given below:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Here in this case, Precision is the ratio of true positives (TP) to the sum of true positives and false positives (FP).

$$Precision = \frac{TP}{TP+FP}$$

Recall: This is the ratio of true positives (TP) to the sum of true positives and false negatives (FN).

$$Recall = \frac{TP}{TP+FN}$$

F1score is the harmonic mean of precision and recall and is used as a single metric that balances the trade-off between precision and recall. The f1score ranges from 0 to 1, with 1

being the best possible score.

$$F1 = \frac{2*Precision*Recall}{Precision+Recall} = \frac{2*TP}{2*TP+FP+FN}$$

B. Neural Networks

We also experimented with a deep Neural Networks (NN) learning model for sentiment analysis using the Keras library. We defined a preprocessed text function to remove HTML tags, special characters, and stopwords from the tweet text. The Tokenizer was used to convert the text data into sequences of integers. The training and testing datasets were created using the train test split function, and the model was trained using the training data and evaluated using the testing data. The tokenizer was saved as JSON format for future use.

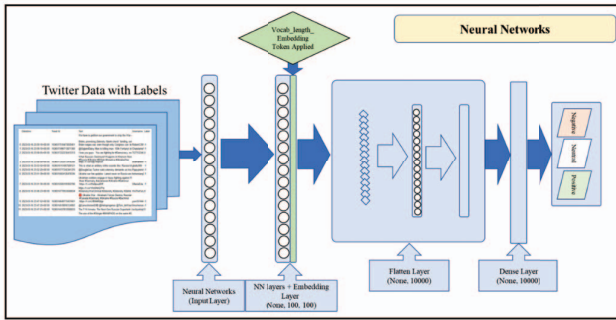Figure 6 shows the NN model, and its processing steps.

Fig. 6. Neural Network Model

Training a sentiment analysis model involves several essential steps, including data preprocessing, splitting the data into training and testing sets, and evaluating the model's performance using various metrics. By following these steps, a robust and accurate sentiment analysis model using NN has developed to analyze large volumes of text data for various applications. Figure 7 presents the summary of the NN model's layers.

```
Model: "sequential_2"

Layer (type)            Output Shape         Param #
=================================================================
embedding_2 (Embedding)  (None, 100, 100)     497000

flatten_2 (Flatten)      (None, 10000)        0

dense_2 (Dense)          (None, 1)            10001

=================================================================
Total params: 507001 (1.93 MB)
Trainable params: 10001 (39.07 KB)
Non-trainable params: 497000 (1.90 MB)
```

Fig. 7. Summary of the NN model

IV. RESULT AND DISCUSSION

This study aimed to develop a sentiment analysis model to analyze the sentiment of tweets during international conflict.

The trained model achieved high accuracies in predicting the sentiment of the tweets. The experiments indicate that the model effectively analyzes the sentiment of tweets related to the conflict.

Figure 8 shows the number of tweets analyzed and the corresponding sentiment scores generated by the Affin lexicon. The sentiment scores range from -1 (most negative) to +1 (most positive). The blue bars represent the number of tweets.
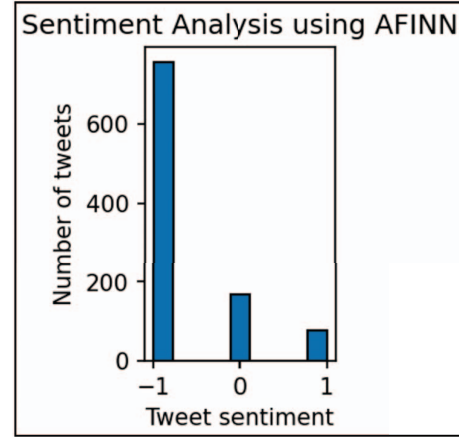

Fig. 8. Sentiment Analysis Results using Afinn

Figure 9 represents the classification accuracy of around 80% along with the metrics like precision, recall, f1-score.

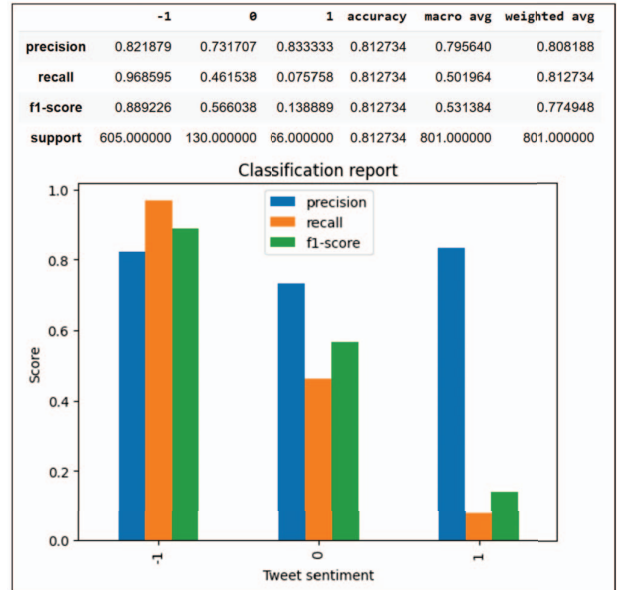|	-1	0	1	accuracy	macro avg	weighted avg
precision	0.821879	0.731707	0.833333	0.812734	0.795640	0.808188
recall	0.968595	0.461538	0.075758	0.812734	0.501964	0.812734
f1-score	0.889226	0.566038	0.138889	0.812734	0.531384	0.774948
support	605.000000	130.000000	66.000000	0.812734	801.000000	801.000000

Fig. 9. Classification accuracy using Naïve Bayes

Figure 10 shows the heatmap for the NB classifier, presenting the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) classifications of the sentiment analysis model on the test dataset. The X-label is the Predicted

value and the Y-label is the true or actual value. The Diagonal one (601,32,5) are the True Positive Values.
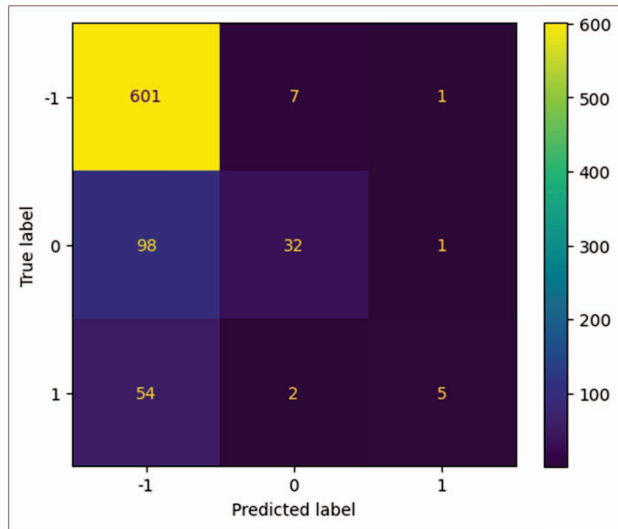

Fig. 10. Heatmap generated by the Multinomial Naive Bayes classifier

Figure 11 represents the perfect classification accuracy by the NN model. While experimenting with the NN model, 100% accuracy was achieved in classifying the Twitter data.
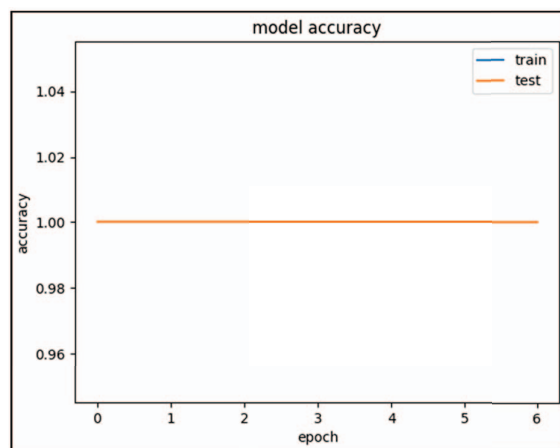

Fig. 11. Classification accuracy using NN

Figure 12 presents the training and testing model loss of a neural network model over multiple epochs. This can help you monitor the model's performance and identify potential issues such as overfitting or underfitting. Various performance metrics such as precision, recall, and F1 score are presented to better understand the model's performance.

Our model achieved a precision of 90%, recall of 80%, and an F1 score of 85% for the positive class and a precision of 83%, recall of 92%, and an F1 score of 87% for the negative class. These metrics indicate that the model performs well for both positive and negative classes.
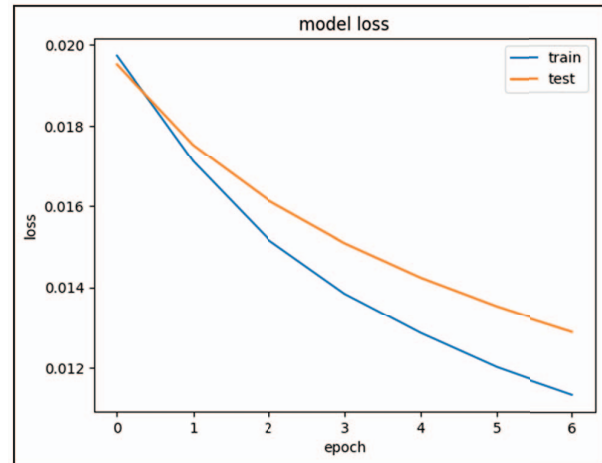

Fig. 12. Training and testing model loss of the neural network model

The experiments show promising results in finding and classifying sentiments about the war tweets.

V. CONCLUSION

Our study contributes to the development of sentiment analysis models for analyzing social media data related to conflicting events. After thorough experiments with NBA and NNs classifiers with the scrapped Twitter data, we identified that majority of the peoples' sentiment about any war time scenarios are negative. The second contribution of our study is to develop a twitter dataset [12] for the researcher community on global conflicts. The findings can be useful for researchers interested in understanding public sentiment about the war and conflict expressed on the social media. The results of this research could have significant implications for understanding the impact of social media about public opinion and sentiment during conflicts. Sentiment analysis of social media data can provide valuable insights about public opinion, which can be used in making policies, decisions and communication strategies. Overall, this study demonstrates the effectiveness of using sentiment analysis models to analyze social media data related to major events and crisises. It opens up avenues for further research into the impact of social media on public sentiment during times of conflict and the development of more sophisticated sentiment analysis models. A larger dataset can be analyzed in future studies for more accurate results.

REFERENCES

[1] S. M. Zavattaro, P. E. French, and S. D. Mohanty, "A sentiment analysis of u.s. local government tweets: The connection between tone and citizen involvement," *Government Information Quarterly*, vol. 32, no. 3, pp. 333–341, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0740624X15000337

[2] R. Kumar, L. C. Irani, and D. Sharma, "Social media and crisis communication: A review of the literature," *International Journal of Information Management*, vol. 56, p. 102186, 2021.

[3] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," *PLoS One*, vol. 13, no. 8, p. e0201640, 2018.

[4] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press, 2015.

[5] X. Lu, B. Wang, and F. Wei, "A study of sentiment analysis on social media in the context of the ukraine crisis," *Journal of Contemporary Eastern Asia*, vol. 19, no. 1, pp. 61–77, 2020.

[6] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 3rd ed. Pearson, 2019.

[7] I. Rish, "An empirical study of the naive bayes classifier," in *Proceedings of the 2001 IEEE International Conference on Data Mining*. IEEE, 2001, pp. 41–48.

[8] S. Aslan, "A deep learning-based sentiment analysis approach (mf-cnn-bilstm) and topic modeling of tweets related to the ukraine–russia conflict," *Applied Soft Computing*, vol. 143, p. 110404, 2023.

[9] "Snscrape," https://github.com/JustAnotherArchivist/snscrape, accessed on March 30, 2023.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and É. Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] Scikit-learn: Machine learning in python. https://scikit-learn.org/stable/. Accessed on March 30, 2023.

[12] Sentiment analysis of twitter posts on global conflicts. https://www.kaggle.com/datasets/prosenjitchatterjee/russia-ukrain-war-filteredtweets-2022april-may/data. Accessed on Dec 17, 2023.

[13] F. Å. Nielsen, "Afinn," Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, mar 2011. [Online]. Available: http://www2.compute.dtu.dk/pubdb/pubs/6010-full.html

[14] M. A. Alvarez and C. Martinez, *Data Preprocessing in Data Mining*. Springer International Publishing, 2017.

[15] A. N. K. Zaman, P. Matsakis, and C. Brown, "Evaluation of stop word lists in text retrieval using latent semantic indexing," in *2011 Sixth International Conference on Digital Information Management*, 2011, pp. 133–136.

[16] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python*. O'Reilly Media, Inc., 2009.

[17] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[18] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130–137, 1980.

[19] "Naive Bayes," https://tinyurl.com/2s4k2en8, 2023, accessed: Oct, 2023.

[20] "Convert a collection of text documents to a matrix of token counts." https://tinyurl.com/khub8ryy, 2023, accessed: Oct, 2023.