Sentiment Analysis on Users Opinion towards the Russia-Ukraine Crisis using Transformers-based Model

Abdulrazaq Ahmed Bello
Department of Computer Science
Federal Polytechnic Bauchi
Bauchi, Nigeria
aabello.cs@fptb.edu.ng

Muhammad Nuruddeen Idris
Department of Computer Science
Federal University Gusau
Zamfara, Nigeria
minuruddeen66@gmail.com

Suberu Yusuf
Department of Computer Science
Federal Polytechnic Bauchi
Bauchi, Nigeria
yusufs@fptb.edu.ng

Ikenna Chukwuka Onuorah
Department of Computer Science
Federal Polytechnic Bauchi
Bauchi, Nigeria
iconourah.cs@fptb.edu.ng

Khalil Sani Ibrahim
Department of Computer Science
Federal Polytechnic Bauchi
Bauchi, Nigeria
ksibrahim.cs@fptb.edu.ng

Lawan Garba
Department of Computer Science
Federal Polytechnice Bauchi
Bauchi, Nigeria
lgarba.cs@fptb.edu.ng

***Abstract*—In this study, we conducted sentiment analysis on Twitter data related to Russia-Ukraine conflict using state-of-the-art transformer-based models, namely DistilBERT and ALBERT. Our dataset spanned the period of two month (a month before the Russia invasion of Ukraine and a month after). The objective was to gauge public sentiment during this period. After preprocessing and fine-tuning, both DistilBERT and ALBERT models achieved remarkable accuracy, with both models scoring 89.9%. The high accuracy suggests that these models are effective in understanding the sentiment expressed in Twitter conversations during the conflict. Our analysis reveals valuable insights into the evolving sentiment of Twitter users during this period showing the most common hashtag in use by users. These findings contribute to a deeper understanding of how social media reflects and shape public opinion during significant geopolitical events. The success of transformer-based models in this context underscores their potential for real-time sentiment analysis and monitoring of public sentiment in response to events happening globally which can be invaluable for policymakers, journalists and researchers seeking to comprehend and respond to rapidly evolving situations.**

Keywords—Sentiment analysis, russia-ukraine crisis, transformer-based model, opinion mining, twitter.

I. INTRODUCTION

Online social networks (OSN) were created to bring people together so they may share their feelings, desires, joy, interests, and hobbies, which is an essential part of socialisation. This is still true, but OSNs have developed into places where people organise themselves to accomplish common goals and where political viewpoints are formed and expressed. The weak moderation and rapid content publication of the most popular platforms set them apart from traditional social media (newspapers, magazines, etc.). As a result, OSNs are crucial in shaping public perception of practically every event. On the other side, they are the perfect medium for disinformation and fake news to spread. Nowadays, social media is a necessary component of most people's everyday lives. Social media and the internet can be used for a variety of things, such as advertising, disseminating financial and political trends, gathering user reviews of items, disseminating spam, and disseminating news [1]. Numerous studies have been conducted on social media from a variety of academic disciplines, including computer science, marketing, psychology, sociology, and marketing. Currently, researchers are analysing tweets using machine learning methods including natural language processing [2], [3].

Natural language processing (NLP) and artificial intelligence (AI) methods for text analytics have received more attention as a result of society's extraordinary increase in its reliance on social media for information as opposed to traditional news sources and the volume of data it presents [4]. This data contains a variety of social phenomena, such as cultural dynamics, social trends, natural disasters, and public health, as well as topics that are widely debated and viewpoints expressed on social media. This is due to the social network's low cost, ease of access, and personal connectivity. Professional opinion leaders (and state actors) are increasingly using social media as a tool to use the network effects of the platform to spread their message. Additionally, a lot of businesses use social media to advertise their goods, brands, and services [5]. Reviews and experiences posted by users produce an information-rich reservoir that is retained as text, making social media and open communication platforms important information sources for researching issues involving fast changing public attitude [6].

NLP has rapidly increased in popularity, and social media analysis using it has seen similar growth. However, it is still difficult to infer a text's intrinsic meaning using NLP-methods. Even the newest NLP tools have been demonstrated to be "vulnerable to adversarial texts." [7], [8].

Recently, a more powerful subset of Machine Learning techniques known as Deep Learning has emerged as a state-of-the-art approach in some developing areas of research [9]. Application domains that have used Deep Learning include Speech Recognition, Computer Vision, and Natural Language Processing. Named Entity Recognition (NER) [10]–[12] and Sentiment Analysis [13], [14] are popular fields of Natural Language Processing (NLP) that have used Deep Learning techniques in this recent times to solve certain complex problems. Deep learning, a kind of machine learning, has only very lately become a potent tool [9] producing state-of-the-art results in many application domains, such as speech, recognition, computer vision, and Natural Language Processing. Strong incentives in this study area have led to several practical applications and industrial interest adopting a variety of methodologies for their diverse sentiment analysis requirements [15]. Earlier approaches have used all types of supervised machine learning methods [16], feature selection [17] methods, and unsupervised machine learning methods [18]. Unsupervised sentiment analysis includes methods that exploit sentiment lexicons, syntactic patterns, and grammatical analysis to classify sentiments [19].

The accuracy of these approaches largely depends on the type of algorithm and the quality of the training data used.

Transformers are a specific architecture within Deep Learning that was introduced to handle sequential data, particularly well-suited for natural language processing tasks. The transformer architecture is not based on recurrent structures like LSTMs. Instead, it employs self-attention mechanisms that allow it to capture relationships between different elements in a sequence, regardless of their distance from each other. Transformers have demonstrated state-of-the-art performance in various NLP tasks and have also been applied to other domains.

Therefore, in this study, Sentiment analysis is conducted on 2022 Russia-Ukraine conflict related tweets using transformer-based model.

II. Related Work

Sentiment analysis is a technique that combines text analysis and Natural Language Processing (NLP) [20] to extract sentiments from reviews to detect favorable and unfavorable opinions towards specific products and services based on a large number of textual data [21].

Over the years, text analysis has been used in various applications: email filtering [22], irony and sarcasm detection [23], document organization [24], sentiment and opinion mining prediction [25], [26], hate speech detection [27][26], question answering systems [28], biomedical text mining [29][25], and more. Twitter data have seen wide usage for emotional analysis [30]–[32]. In other work, more than 70 000 tweets made over a year were used in a study analyzing customer feedback of a French Energy Company [33].

The analysis of past epidemics, crisis situational analysis and tracking, has also involved tracking Twitter data [34]–[36]. There has been a huge advancement of what has become web 2.0 today and its related technologies [37]. This advancement has brought about a constant increase in the number of users accessing the web along with a competitive rise in the creation of online services and content [38]. Some of the innovations have enabled the meeting and administration of a large universal crowd with common interest through social interaction through websites that emphasize user-generated content, participatory culture, and interoperability on easily accessible platforms. This is facilitated through social networks available on the web such as Facebook, Twitter, Pinterest, Reddit, YouTube and LinkedIn.

Twitter is one of the most popular social media sites. Users post 500 million tweets daily in average on Twitter [39] and 80% of those tweets are posted from a mobile device [40]. Furthermore, the number of active Twitter users exceeds 22% of the internet users in the world [41]. As of December 2022, Twitter's audience accounted for over 368 million monthly active users worldwide.

The occurrence of social crises nowadays is an influential concern and arduous to eliminate instantly. Crisis in social life becomes a concern of the government and people of a community [42]. Every situation that becomes a reason of danger for social life, people consider it as a social crisis. If it can be identified, bad effects can be minimized and the problem can be eliminated. During any social crisis, reliable information is essential for solving that issue. Many organizations have already utilized the power of social media but a few teams worked for the people in social crisis [43].

According to the work of [44], the internet has become an increasingly important tool in the political realm, both for enabling politicians to reach large numbers of people quickly, and for managing crisis communication. Politicians have increasingly been turning to the internet as a way to get their message out, as well as to manage their reputation in the event of a political crisis [45]. Moreover, politicians can use social media to connect with their constituents and explain their policies and positions to them. They can also use social media to quickly respond to any criticism or questions from their constituents [46]. In addition to using social media, politicians also make use of websites to disseminate their message. Websites can provide a more in-depth look at a politician’s views and policy positions. They also serve as a central hub for all of the information related to a politician’s campaign [47]. In the event of a political crisis, the internet can be a powerful tool for politicians to manage their reputation. Politicians can use the internet to quickly respond to public criticism, address any inaccuracies in the press, and explain their positions in a clear and meaningful way [48]. By being proactive and providing accurate information, politicians can help limit the damage of a political crisis. The role of the internet in disseminating the message of politicians and managing crisis communication is to provide a platform for politicians, journalists, and citizens to interact and share information quickly and efficiently [49].

The sentiment indicators of leaders can be measured through the discourse analysis of the various messages posted across various social media platforms by these leaders [50]. This

discourse analysis can be done through methods such as textual analysis of their posts, sentiments expressed by other users in response to their posts, and sentiment ratings given by independent researchers. Textual analysis of posts includes analysing the words and phrases used, understanding how the language is used to communicate certain ideas, and any emotional context or tone that are present in the post [51].

In another study, [52] investigated the public opinions and sentiments towards Syrian refugees using text data from Twitter. The author also stated their work to be the first sentiment analysis study that explores the sentiments towards Syrian civil war and refugee crisis in a comparative analysis using data from social media. Social media is considered as a good venue because a mass number of people from various backgrounds use social media to discuss the topic and people are less reluctant to share their sincere thoughts on social media due to anonymity.

In another study, [53] Investigated the Twitter usage behaviors of journalists. The tsunami warnings in Padang Indonesia and reactions among Twitter users have been examined [54]. A study by [39] investigated Twitter usage behaviors of cancer patients. They found that cancer patients describe and explain their feelings about their diseases openly and candidly on Twitter.

Based on 36 million tweets collected from Twitter, [2] proposed a real-time sentiment analysis system for classification of political tweets during 2012 US presidential elections. Their model achieved 59% accuracy in predicting the sentiments of political tweets.

In another research study, [55] investigated the propagation of the issues discussed during TV debates in the 2012 South Korean presidential election to conversations on Twitter and communication patterns of Twitter users in political discussions. In another related work, [56] proposed an approach to predict the movement directions in Dow Jones Industrial Average closing values. Their method provided 86.7% accuracy in predicting the daily up and down directions of Dow Jones closing values.

On the other hand, [57] offered a sentiment analysis model that provides useful graphical visualizations about potential terrorism scenarios based on public sentiment data collected on Twitter. In a case study, [58] have explored the use social media technologies and their impacts on knowledge sharing and effective decision-making during 2010 Haitian Earthquake. The case study demonstrated an example that social media and knowledge management technologies can play a critical role in effective response to disasters by sharing emergency information. In another recent study, [59] investigated the characteristics of social networks after a disaster by investigating the social media responses after the 2016 flood in Louisiana. They observed that the disaster related information diffused differently on social networks during emergency situations.

From text classification perspective, [60] performed sentiment analysis on 30,000 Chinese microblogs and they classified messages into negative, neutral and positive categories. Using movie data, [61] examined the sentiments of 50,000 movie reviews. [62] also perform a study in identifying Fake News in the Russian-Ukrainian Conflict Using Machine Learning. The results of the study show that it is possible to develop a system that can discern between real and fake news regarding the Russian-Ukrainian conflict. Support vector machine and logistic regression outperform other learning algorithms and produce comparable prediction models with approximately 76% of accuracy. [63] perform a Comprehensive Evaluation of Supervised Machine Learning Models Using Twitter Data on Russia–Ukraine War. Various model were assessed using a number of important performance indicators, including accuracy, precision, recall, and F1 score. The results show that the extra trees classifier (ETC) model achieves a highest accuracy of 0.84 in combination with the Bow property which is a measure to evaluate the efficacy of a machine learning algorithm. Logistic regression (LR), decision tree (DT), support vector machine (SVM), XGB, Gaussian naive base (GNB), ADA, and K-nearest neighbours (KNN) comparison were also made.

III. Methodology

A. Data Collection

For this research, data will be extracted from Twitter for experiments. The choice of Twitter is as a result of the simplicity it offers through the use of the Twitter Application Programming Interface (API) for the extraction of hsitorical tweets as opposed to using web scraping techniques. With the Twitter API, we will search for the needed tweets by using a particular word or hashtag. We will use Tweepy, an official Python Twitter API library to generate the datasets.

Over a million tweets were crawled from January to March 2022. The datasets contains two months' worth of tweets. The start date of the tweets search is a month before the Russia invasion and a month after the invasion. (23rd January 2022 to 24th March 2022). The search is limited to tweets in English. The keywords to be use ensure a textual corpus focused on the Russia-Ukraine conflict and associated phenomena. Some of the Keywords to be use to collect the tweets are given in Table 1.

B. Labeling

In order to annotate the data set, we followed guidelines by [64] and labeled each tweet as positive, negative, or neutral. The TextBlob tool was used for the purpose of labeling the emotional sentiment into positive, negative, and neutral. According to [1], TextBlob can indicate a sentence's attitude by calculating the score as a polarity [-1 to 1]. When the polarity of a tweet is less than −0.4, its sentiment is regarded as negative. Symmetrically, when it is greater than 0.4, the tweet is considered positive. The polarity of a neutral tweet lies in the range from −0.4 to 0.4. The pseudocodes of tweets labeling and sentiment are given in Algorithm 1

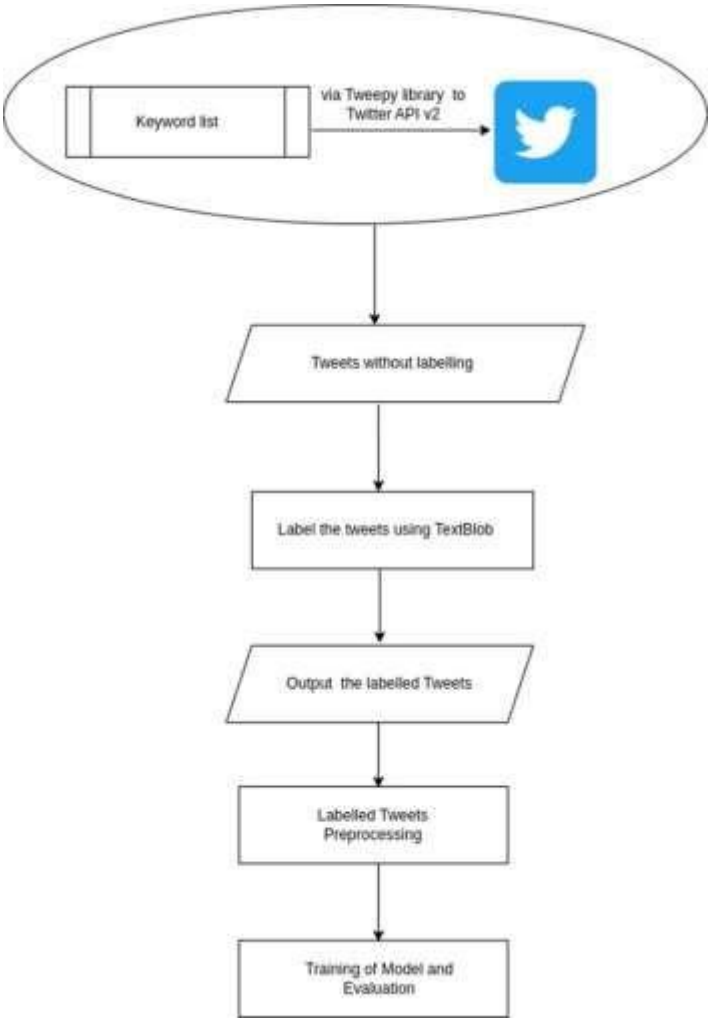

Figure 1: Overview of methodology

Table 1: Search keyword

Keywords use to collect tweets
Zelenskyy
Putin
War
EU
NATO
Russia
Ukraine
RussiaUkraine
RussiaUkraineWar

Algorithm 1 Tweets Labeling Steps

Input: Unlabeled Tweet: T_u,

Output: Labeled Tweet: T_l,

Compute:

Positive: T_{pos} = [];

Negative: T_{neg} = [];

Neutral: T_{neu} = [];

Steps:

for t in T(t) do:

If (t is English):

Perform Labeling: (TextBlob)

If (-1 < polarity of t < -0.4):

Labeled as Negative

else If (0.4 < polarity of t < -0.4):

Labeled as Positive

else:

Labeled as Neutral

Perform Pre-processing:

Remove punctuation, stop-words, stemming and lower-case all words

else:

Delete t

end for

Output:

Pre-processed labeled Tweets: $T_l = [T_{pos}, T_{neg}, T_{neu}]$

Algorithm 1 (cited from [25])

C. Data Preprocessing

As raw tweets are often short, unstructured, informal, and noisy, the first step of sentiment analysis is to preprocess the data. To do so, the following series of techniques are applied in the given order to improve the text [25].

1) Almost every social media platform uses hashtags to represent topics, i.e., #War, #Russia #Ukraine and #NATO. In most cases, hashtags are unnecessary to sentiment and can affect the performance. Thus, in our first step, we will perform basic cleaning of the text by removing unnecessary hashtags, just the hashtag character not the hashtag text.

2) The second step is to case-fold the text. To avoid recognizing the same word as a different word due to capitalization, we will fold all capitalized letters to lower case.

3) There are many words that are concatenated in other words, especially the words that are hashtagged such as “stoptheinvasion” and “standwithukraine,” should be “stop the invasion” and “stand with ukraine,” respectively. Hence, we will perform word segmentation to achieve this.

4) Removing stop words is a popular method to reduce the noise in textual data. Removing stop words does not affect understanding a sentence’s sentiment valence. For all but sentiment analysis, we will remove stop words (words commonly found in a document of little semantic value, e.g., “for,” “the,” and “is”).

5) The fifth step is lemmatization that processes the morphological analysis of words and returns words to the base or dictionary form. We will use the method provided in the nltk (Natural Language ToolKit) and lemmatize words (changing

different forms of a word to its basic root form, e.g., “viruses” to “virus” or “went” to “go”).

6) Textual data analysis will be done to eliminate hyperlinks, @mentions, and punctuation. We will remove special characters, punctuation, and numbers from the data set as these do not help detect sentiment.

D. *Training and Evaluation*

Transformer-based models [7], such as bidirectional encoder representations from the transformer (BERT) [22] and a distilled version of BERT (DistilBERT) [79], which reduces the size of a BERT LM by 40% while retaining 97% of its language understanding proficiency and is 60% quicker, are current state-of-the-art language models (LMs) that have altered NLP research’s landscape [57]. A Lite Bidirectional Encoder Representations from Transformers (ALBERT) [91] that is modified based on the architecture of BERT in scaling pretrained models. ALBERT implements two-parameter reduction methods that lift the essential barriers: 1) factorize embedding parameterization—decomposing a big vocabulary embedding matrix to two small matrices; 2) replace the next sentence prediction (NSP) loss by sentence order prediction (SOP) loss; and 3) provide cross-layer parameter sharing, which are trained using bidirectional transformer models on large-scale unlabeled text corpus using mask language modeling (MLM), NSP, and SOP algorithms. These transformer-based models differ in how the training methods are used. However, generally, all training is performed in an unsupervised manner to generate an LM used as an input to the subsequent supervised fine-tuning for NLP-tasks.

Evaluation metrics quantify the performance of a predictive model [75]. A widely used case is the model classification accuracy. For this study, we will be using only the Accuracy metric for evaluation.

Mathematically, we have,

$$Accuracyofamodel = \frac{(TP + TN)}{(TP + FN + FP + TN)}$$

IV. Results and Discussion

An extensive dataset of 1,592,350 tweets were analyzed in this study. They provide a robust foundation for understanding public sentiment surrounding the Russia-Ukraine conflict. This large volume of data shows that there were diverse range of opinions expressed on Twitter during this critical period. From the total tweets, 1,432,775 were labeled as neutral. 96,753 were labeled as positive and 62,822 were labeled as negative. Fig 2 shows the distribution of the top ten hashtag. From the study, 1,106,986 hashtag were use during the period of data collection. 94,967 represent the number of repeated hashtag. The graphical representation of the labeled tweets is shown in Fig 3.

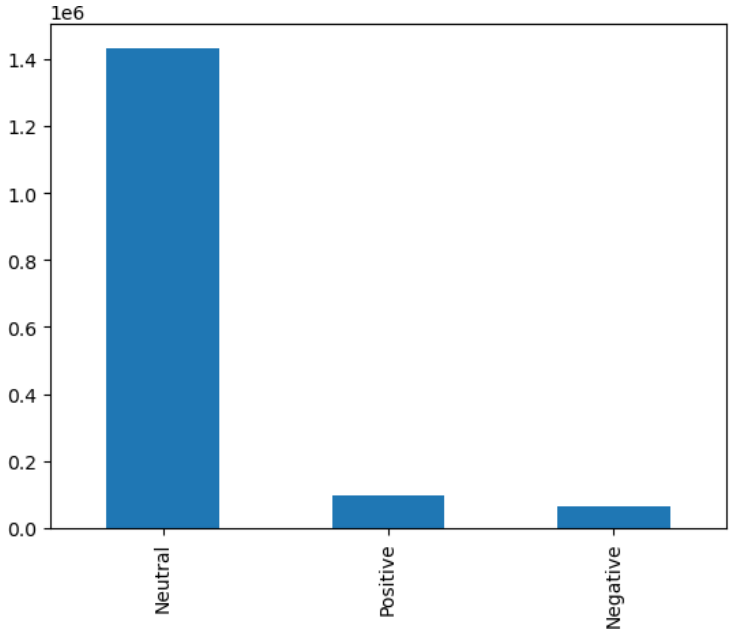

Figure 3: Distribution of labeled tweets

Due to computational resources, the model was trained for only one epoch and they both yield an accuracy of 89.9%.

A. *Imbalanced Sentiment Distribution*

One notable aspect of the dataset is the substantial class imbalance. The majority of tweets (approximately 90%) are labeled as ‘neutral’, while ‘positive’ and ‘negative’ sentiments account for around 6% and 4% respectively. This imbalance raises important considerations regarding the representativeness of the sentiment labels. It suggests that, in the context of this conflict, a significant portion of the public tweets on Twitter may be characterized by a relatively objective or non-committal tone.

B. *Limitations and Future Directions*

While this study provides valuable insights, it is not without limitations. The analysis is limited to Twitter data and may not capture sentiment expressed in other platforms (such as Reddit, Quora) or offline. Additionally, the sentiment labels are generated based on automated methods and may not always capture context-dependent expression accurately. Future research could delve deeper into the content of tweets to identify key themes and topics that elicit particular sentiments. Moreover, exploring the temporal evolution of sentiment before, during and after critical events offers a more understanding of public reactions over time.

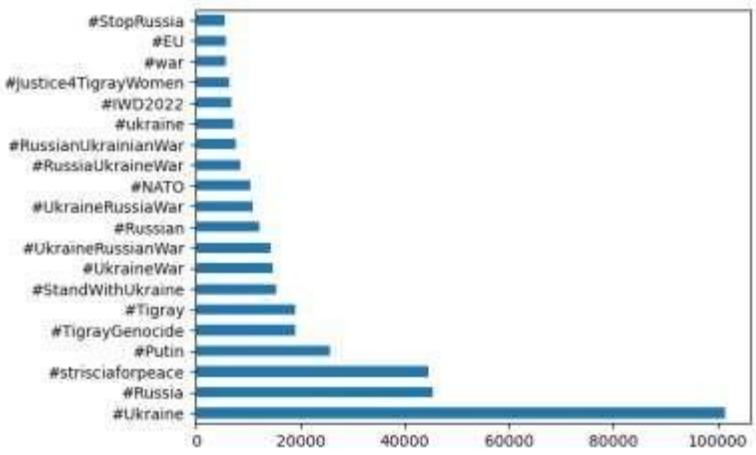

Figure 2: Top ten hashtag

V. CONCLUSION

In conclusion, our study not only advances our understanding of sentiment dynamics in the context of the Russia-Ukraine conflict but also highlights the substantial potential of advanced natural language processing techniques in capturing and analyzing public sentiment during times of geopolitical significance. It is also expected going forward that the limitation of imbalance sentiment distribution in the datasets experienced in this study would be dived into by other researchers, thereby providing a valuable foundation for future research in the field of sentiment analysis and its application in understanding of global events.

REFERENCES

[1] A. Alarifi, M. Alsaleh, and A. M. Al-Salman, "Twitter turing test: Identifying social machines," Inf Sci (N Y), vol. 372, pp. 332–346, Dec. 2016, doi: 10.1016/J.INS.2016.08.036.

[2] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle," in Proceedings of the ACL 2012 System Demonstrations, Jeju Island, Korea: Association for Computational Linguistics, Jul. 2012, pp. 115–120. [Online]. Available: https://aclanthology.org/P12-3020

[3] L. Schweitzer, "Planning and Social Media: A Case Study of Public Transit and Stigma on Twitter," Journal of the American Planning Association, vol. 80, no. 3, pp. 218–238, 2014, doi: 10.1080/01944363.2014.980439.

[4] F. Barbieri and H. Saggion, "Automatic Detection of Irony and Humour in Twitter," in International Conference on Innovative Computing and Cloud Computing, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:2139885

[5] W. He, H. Wu, G. Yan, V. Akula, and J. Shen, "A novel social media competitive analytics framework with sentiment benchmarks," Information & Management, vol. 52, no. 7, pp. 801–812, 2015, doi: https://doi.org/10.1016/j.im.2015.04.006.

[6] M. J. Widener and W. Li, "Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US," Applied Geography, vol. 54, pp. 189–197, 2014, doi: https://doi.org/10.1016/j.apgeog.2014.07.017.

[7] Z. Saeed, R. A. Abbasi, I. Razzak, O. Maqbool, A. Sadaf, and G. Xu, "Enhanced Heartbeat Graph for emerging event detection on Twitter using time series networks," Expert Syst Appl, vol. 136, pp. 115–132, 2019, doi: https://doi.org/10.1016/j.eswa.2019.06.005.

[8] Z. Saeed et al., "What's Happening Around the World? A Survey and Framework on Event Detection Techniques on Twitter," J Grid Comput, vol. 17, no. 2, pp. 279–312, 2019, doi: 10.1007/s10723-019-09482-2.

[9] J. Heaton, "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning," Genet Program Evolvable Mach, vol. 19, no. 1, pp. 305–307, 2018, doi: 10.1007/s10710-017-9314-z.

[10] J. Liu et al., "A hybrid deep-learning approach for complex biochemical named entity recognition," CoRR, vol. abs/2012.10824, 2020, [Online]. Available: https://arxiv.org/abs/2012.10824

[11] Q. Zhang, Y. Sun, L. Zhang, Y. Jiao, and Y. Tian, "Named entity recognition method in health preserving field based on BERT," Procedia Comput Sci, vol. 183, pp. 212–220, 2021, doi: https://doi.org/10.1016/j.procs.2021.03.010.

[12] T. Mehmood, A. E. Gerevini, A. Lavelli, and I. Serina, "Combining Multi-task Learning with Transfer Learning for Biomedical Named Entity Recognition," Procedia Comput Sci, vol. 176, pp. 848–857, 2020, doi: https://doi.org/10.1016/j.procs.2020.09.080.

[13] A. Baliyan, A. Batra, and S. P. Singh, "Multilingual Sentiment Analysis using RNN-LSTM and Neural Machine Translation," in 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), 2021, pp. 710–713.

[14] N. Hossain, Md. R. Bhuiyan, Z. N. Tumpa, and S. A. Hossain, "Sentiment Analysis of Restaurant Reviews using Combined CNN-LSTM," in 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1–5. doi: 10.1109/ICCCNT49239.2020.9225328.

[15] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," Sep. 2019.

[16] H. M. K. Kumar and B. S. Harish, "A New Feature Selection Method for Sentiment Analysis in Short Text," Journal of Intelligent Systems, vol. 29, no. 1, pp. 1122–1134, Dec. 2019, doi: 10.1515/jisys-2018-0171.

[17] A. Madasu and S. Elango, "Efficient feature selection techniques for sentiment analysis," Multimed Tools Appl, vol. 79, no. 9, pp. 6313–6335, 2020, doi: 10.1007/s11042-019-08409-z.

[18] A. Chandra Pandey, D. Singh Rajpoot, and M. Saraswat, "Twitter sentiment analysis using hybrid cuckoo search method," Inf Process Manag, vol. 53, no. 4, pp. 764–779, Jul. 2017, doi: 10.1016/j.ipm.2017.02.004.

[19] M. Z. Asghar, A. Khan, S. Ahmad, M. Qasim, and I. A. Khan, "Lexicon-enhanced sentiment analysis framework using rule-based classification scheme," PLoS One, vol. 12, no. 2, p. e0171649, Feb. 2017, doi: 10.1371/journal.pone.0171649.

[20] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques," in Third IEEE International Conference on Data Mining, 2003, pp. 427–434. doi: 10.1109/ICDM.2003.1250949.

[21] D. A. Al-Qudah, A. M. Al-Zoubi, P. A. Castillo-Valdivieso, and H. Faris, "Sentiment Analysis for e-Payment Service Providers Using Evolutionary eXtreme Gradient Boosting," IEEE Access, vol. 8, pp. 189930–189944, 2020, doi: 10.1109/ACCESS.2020.3032216.

[22] X. Carreras and L. Màrquez, "Boosting Trees for Anti-Spam Email Filtering," CoRR, vol. cs.CL/0109015, 2001, [Online]. Available: https://arxiv.org/abs/cs/0109015

[23] U. Naseem, I. Razzak, P. Eklund, and K. Musial, "Towards Improved Deep Contextual Embedding for the identification of Irony and Sarcasm," in 2020 International Joint Conference on Neural Networks (IJCNN), IEEE, Jul. 2020, pp. 1–7. doi: 10.1109/IJCNN48605.2020.9207237.

[24] K. M. Hammouda and M. S. Kamel, "Efficient phrase-based document indexing for Web document clustering," IEEE Trans Knowl Data Eng, vol. 16, no. 10, pp. 1279–1296, Oct. 2004, doi: 10.1109/TKDE.2004.58.

[25] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund, and J. Kim, "COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis," IEEE Trans Comput Soc Syst, vol. 8, no. 4, pp. 1003–1015, Aug. 2021, doi: 10.1109/TCSS.2021.3051189.

[26] U. Naseem, S. K. Khan, I. Razzak, and I. A. Hameed, "Hybrid Words Representation for Airlines Sentiment Analysis," 2019, pp. 381–392. doi: 10.1007/978-3-030-35288-2_31.

[27] U. Naseem, S. K. Khan, F. Madiha, and F. Ali, “Abusive Language Detection: A Comprehensive Review,” Indian J Sci Technol, vol. 12, no. 45, pp. 01–13, Dec. 2019, doi: 10.17485/ijst/2019/v12i45/146538.

[28] V. Gupta and G. S. Lehal, “A Survey of Text Mining Techniques and Applications,” Journal of Emerging Technologies in Web Intelligence, vol. 1, no. 1, Aug. 2009, doi: 10.4304/jetwi.1.1.60-76.

[29] U. Naseem, I. Razzak, and P. W. Eklund, “A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter,” Multimed Tools Appl, vol. 80, no. 28–29, pp. 35239–35266, Nov. 2021, doi: 10.1007/s11042-020-10082-6.

[30] G. Rocha and H. Lopes Cardoso, “Recognizing Textual Entailment: Challenges in the Portuguese Language,” Information, vol. 9, no. 4, p. 76, Mar. 2018, doi: 10.3390/info9040076.

[31] G. Carducci, G. Rizzo, D. Monti, E. Palumbo, and M. Morisio, “TwitPersonality: Computing Personality Traits from Tweets Using Word Embeddings and Supervised Learning,” Information, vol. 9, no. 5, 2018, doi: 10.3390/info9050127.

[32] T. Ahmad, A. Ramsay, and H. Ahmed, “Detecting Emotions in English and Arabic Tweets,” Information, vol. 10, no. 3, p. 98, Mar. 2019, doi: 10.3390/info10030098.

[33] L. Pépin, P. Kuntz, J. Blanchard, F. Guillet, and P. Suignard, “Visual analytics for exploring topic long-term evolution and detecting weak signals in company targeted tweets,” Comput Ind Eng, vol. 112, pp. 450–458, Oct. 2017, doi: 10.1016/j.cie.2017.01.025.

[34] X. Ye, S. Li, X. Yang, and C. Qin, “Use of Social Media for the Detection and Analysis of Infectious Diseases in China,” ISPRS Int J Geoinf, vol. 5, no. 9, p. 156, Aug. 2016, doi: 10.3390/ijgi5090156.

[35] J. Kim and M. Hastak, “Social network analysis: Characteristics of online social networks after a disaster,” Int J Inf Manage, vol. 38, no. 1, pp. 86–96, Feb. 2018, doi: 10.1016/j.ijinfomgt.2017.08.003.

[36] I. Fung et al., “Pedagogical Demonstration of Twitter Data Analysis: A Case Study of World AIDS Day, 2014,” Data (Basel), vol. 4, no. 2, p. 84, Jun. 2019, doi: 10.3390/data4020084.

[37] K. Patel, “Incremental Journey for World Wide Web: Introduced with Web 1.0 to Recent Web 5.0 - A Survey Paper,” 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:16012196

[38] T. O’reilly, “What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software.”

[39] W. C. Crannell, E. Clark, C. Jones, T. A. James, and J. Moore, “A pattern-matched Twitter analysis of US cancer-patient sentiments,” Journal of Surgical Research, vol. 206, no. 2, pp. 536–542, 2016, doi: https://doi.org/10.1016/j.jss.2016.06.050.

[40] [40] K. M. Carley, M. Malik, P. M. Landwehr, J. Pfeffer, and M. Kowalchuck, “Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia,” Saf Sci, vol. 90, pp. 48–61, 2016, doi: https://doi.org/10.1016/j.ssci.2016.04.002.

[41] V. Kayser and A. Bierwisch, “Using Twitter for foresight: An opportunity?,” Futures, vol. 84, pp. 50–63, Nov. 2016, doi: 10.1016/j.futures.2016.09.006.

[42] P. Panagiotopoulos, J. Barnett, A. Z. Bigdeli, and S. Sams, “Social media in emergency management: Twitter as a tool for communicating risks to the public,” Technol Forecast Soc Change, vol. 111, pp. 86–96, Oct. 2016, doi: 10.1016/j.techfore.2016.06.010.

[43] L. Austin and Y. Jin, Eds., Social Media and Crisis Communication. Routledge, 2017. doi: 10.4324/9781315749068.

[44] “SSRN-id4320303”.

[45] S. Splendiani and A. Capriello, “Crisis communication, social media and natural disasters – the use of Twitter by local governments during the 2016 Italian earthquake,” Corporate Communications: An International Journal, vol. 27, no. 3, pp. 509–526, Jun. 2022, doi: 10.1108/CCIJ-03-2021-0036.

[46] C. White, Social Media, Crisis Communication, and Emergency Management: Leveraging Web 2.0 Technologies. 2011. doi: 10.1201/b11251.

[47] M. Barisione and A. Michailidou, “Do We Need to Rethink EU Politics in the Social Media Era? An Introduction to the Volume,” in Social Media and European Politics, London: Palgrave Macmillan UK, 2017, pp. 1–23. doi: 10.1057/978-1-137-59890-5_1.

[48] Y. Cheng, “How Social Media Is Changing Crisis Communication Strategies: Evidence from the Updated Literature,” Journal of Contingencies and Crisis Management, vol. 26, no. 1, pp. 58–68, Mar. 2018, doi: 10.1111/1468-5973.12130.

[49] D. Nemer and M. Tsikerdekis, “Political engagement and ICTs: Internet use in marginalized communities,” J Assoc Inf Sci Technol, vol. 68, no. 6, pp. 1539–1550, Jun. 2017, doi: 10.1002/asi.23779.

[50] L. Oliveira and Á. Figueira, “Social Media Content Analysis in the Higher Education Sector,” International Journal of Web Portals, vol. 7, no. 2, pp. 16–32, Apr. 2015, doi: 10.4018/IJWP.2015040102.

[51] M. Mameli, M. Paolanti, C. Morbidoni, E. Frontoni, and A. Teti, “Social media analytics system for action inspection on social networks,” Soc Netw Anal Min, vol. 12, no. 1, p. 33, Dec. 2022, doi: 10.1007/s13278-021-00853-w.

[52] 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, Apr. 2018, doi: 10.1016/j.tele.2017.10.006.

[53] N. Y. Lee, Y. Kim, and Y. Sang, “How do journalists leverage Twitter? Expressive and consumptive use of Twitter,” Soc Sci J, vol. 54, no. 2, pp. 139–147, Jun. 2017, doi: 10.1016/j.soscij.2016.09.004.

[54] K. M. Carley, M. Malik, P. M. Landwehr, J. Pfeffer, and M. Kowalchuck, “Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia,” Saf Sci, vol. 90, pp. 48–61, 2016, doi: https://doi.org/10.1016/j.ssci.2016.04.002.

[55] S. J. Park, J. Y. Park, Y. S. Lim, and H. W. Park, “Expanding the presidential debate by tweeting: The 2012 presidential election debate in South Korea,” Telematics and Informatics, vol. 33, no. 2, pp. 557–569, May 2016, doi: 10.1016/j.tele.2015.08.004.

[56] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” J Comput Sci, vol. 2, no. 1, pp. 1–8, Mar. 2011, doi: 10.1016/j.jocs.2010.12.007.

[57] M. Cheong and V. C. S. Lee, “A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter,” Information Systems Frontiers, vol. 13, no. 1, pp. 45–59, Mar. 2011, doi: 10.1007/s10796-010-9273-x.

[58] D. Yates and S. Paquette, “Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake,” Int J Inf Manage, vol. 31, no. 1, pp. 6–13, 2011, doi: https://doi.org/10.1016/j.ijinfomgt.2010.10.001.

[59] J. Kim and M. Hastak, “Social network analysis: Characteristics of online social networks after a disaster,” Int J Inf Manage, vol. 38, no. 1, pp. 86–96, Feb. 2018, doi: 10.1016/j.ijinfomgt.2017.08.003.

[60] C. Liao, C. Feng, S. Yang, and H. Huang, “Topic-related Chinese message sentiment analysis,” Neurocomputing, vol. 210, pp. 237–246, Oct. 2016, doi: 10.1016/j.neucom.2016.01.110.

[61] F. H. Khan, U. Qamar, and S. Bashir, “eSAP: A decision support framework for enhanced sentiment analysis and polarity classification,” Inf Sci (N Y), vol. 367–368, pp. 862–873, Nov. 2016, doi: 10.1016/j.ins.2016.07.028.

[62] Y. and M. M. and A. R. and A. R. and A. A. and A. M. and E. M. Darwish Omar and Tashtoush, “Identifying Fake News in the Russian-Ukrainian Conflict Using Machine Learning,” in Advanced Information Networking and Applications, L. Barolli, Ed., Cham: Springer International Publishing, 2023, pp. 546–557.

[63] G. K. Wadhwani, P. K. Varshney, A. Gupta, and S. Kumar, “Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models Using Twitter Data on Russia–Ukraine War,” SN Comput Sci, vol. 4, no. 4, p. 346, Apr. 2023, doi: 10.1007/s42979-023-01790-5.

[64] A. Bandi and A. Fellah, “Socio-Analyzer: A Sentiment Analysis Using Social Media Data,” pp. 61–53. doi: 10.29007/kzk1.