Stance Classification on FIFA World Cup Using Twitter Data

Aaquib Asrar, Susmita Das, and Sangita Dutta

Abstract The widespread influence of social media has led to an increase in online discussions on varied topics. International sports events with a worldwide following bring in a lot of online engagement between people with various viewpoints. Analyzing the stances in different public discourses from social media posts has gained research attraction. In this paper, we have tried to identify the stance of people on the FIFA World Cup 2022 from Twitter. Natural language processing techniques have been used for text preprocessing and compared different machine learning approaches to classify the Twitter text content into three stance classes of FAVOR, AGAINST, and NONE. We have observed that for our approach, support vector machine (SVM) provides the best result with good accuracy. Our approach is performing well on benchmark datasets.

Keywords Stance detection · Machine learning · Natural language processing

1 Introduction

In this period of the Internet, social media platforms are becoming dominant and replacing traditional online communication methods. An increasing number of people are obtaining news and information from social media. Sometimes, these platforms play a crucial role in determining public discourse regarding different issues and concepts. Researchers are gaining interest in understanding these online conversations and ascertaining the stance of the users on a certain topic from their posts on social media.

A. Asrar · S. Das · S. Dutta (✉)
Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Howrah 711103, India
e-mail: duttasangita72@gmail.com

A. Asrar
e-mail: 2020csb058.aaquib@students.iiests.ac.in

S. Das
e-mail: susmitad900@gmail.com

A. K. Das et al. (eds.), *Computational Intelligence in Pattern Recognition*,
Lecture Notes in Networks and Systems 725,
https://doi.org/10.1007/978-981-99-3734-9_18

We have considered the recently concluded FIFA World Cup 2022 held in Qatar for our research. It has attracted spectators from all over the world and has one of the highest viewership and ratings. This global event became a primary subject of discussion on social media platforms. There are previous works on analyzing the bias [11] and sentiment of tweets [8] specific to Twitter. Detection of stances from social media regarding football has not been provided much attention.

In our work, the stances of viewers toward the tournament have been analyzed as a whole using NLP and machine learning approaches. The rest of the paper contains related works in Sect. 2, proposed method in Sect. 3, result and analysis in Sect. 4, and conclusion in Sect. 5.

2 Related Works

Stance detection has been a trending area of research over the past few years. Dey et al. [5] have conducted the classification of tweets as neutral and non-neutral based on subjectivity. Hou et al. [10] studied the detection of stance with regard to COVID misinformation on Twitter. Grimminger et al. [6] examined political rivalry and general public stance detection using political tweets. Das et al. [4] analyzed the perception of people on climate change issues and the conference of parties(COP).

Haouari et al. [7] established the aim of identifying authorities' stance about rumors in tweets and made the first dataset available for the work focusing on rumors in Arabic. They investigated the efficacy of Arabic datasets already available for stance detection. Cao et al. [3] presented the overview of stance detection. Essential steps of the framework are described in depth as they first present a broad framework for stance detection. The three groups of cutting-edge stance detection techniques are feature-based techniques, deep learning techniques, and ensemble learning techniques. Also, the benefits and drawbacks of the current methodologies are examined. According to the survey results, hybrid neural network-based solutions are more effective than other methods. Price et al. [13] have researched the influence of Twitter in molding the relationship of different football clubs with their supporters. Hidayatullah et al. [9] conducted topic modeling to analyze the topics that are discussed in social media platforms regarding football news. Buongiovanni et al. [1] analyzed the creation of echo chambers on Twitter during Euro 2020 tournament in Italy regarding a specific topic. Burgers et al. [2] have also studied the bias in social media regarding football in live commentaries.

We have studied the interactions on Twitter during the whole FIFA World Cup 2022 Tournament and observed the stances of viewers.

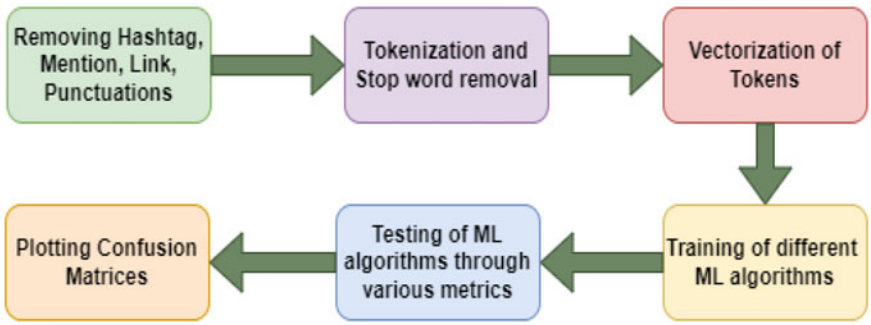


Fig. 1 Framework of proposed method

3 Proposed Method

We have proposed our novel method for stance classification using both natural language processing and machine learning methods.

3.1 Data Collection and Labeling

We considered Twitter for our research, as most of the discussions between different individuals with a variety of opinions are exchanged on Twitter. Tweets were extracted using the Twitter API for the hashtag #FIFAWorldCup2022. A total of 15, 372 tweets from 3rd December to 18th December 2022 have been extracted. Stances of all the tweets were labeled accordingly into three main categories: FAVOR, AGAINST, and NONE. To represent these stances in numbers, 1 for FAVOR, 0 or NONE, and -1 for AGAINST have been considered. In labeling their stances, the target was to understand the awareness and the viewpoints regarding FIFA World Cup 2022 on social media platforms. The tweet posts written in English have been considered, tweet posts in any other language were removed. The tweets have been vetted and considered the ones relevant to the tournament. The tweets with scoreline of any match have been removed as it would not make much sense in deciding the stance of the tweet. The framework of the proposed method is shown in Fig. 1.

3.2 Preprocessing

The raw tweet data becomes difficult to handle, and the basic functions of natural language processing have been utilized for preprocessing. We tokenized the tweet text, removed all the stopwords and punctuations from the tweets. Emoticons have also been considered as they sometimes become significant in representing the stance of a particular tweet. The pronouns are converted to lowercase as it doesn't create much difference in meaning, but as an example, "the" and "The" which are not pronouns signify different meanings. All the user mentions, hashtags, special characters, and URLs in the tweets were removed because they have no significance in deciding the stance of a tweet. Misspelt words were corrected and words that were in contracted form were normalized for ease of computation. The corresponding modules in NLP tools such as TweetNLP, Stanford CoreNLP, and NLTK were used for our preprocessing purposes.

3.3 Model Selection

In the proposed model, the preprocessed text data is taken as input, and the words or tokens are converted into vectors using the TF-IDF vectorizer.

TF-IDF Vectorizer: It is a vectorizer that combines the concepts of term frequency(TF) depicted in Eq. (1) and inverse document frequency(IDF) depicted in Eq. (2) to transform the text into vectors. TF is formulated as

$$\text{Term Frequency}_x(\text{TF}_x) = \frac{T_x}{F} \tag{1}$$

where T_x denotes the number of times the term x appears in the document and F denotes total number of terms in the document. IDF is formulated as

$$\text{Inverse Document Frequency}_x(\text{IDF}_x) = \log\left(\frac{N}{\text{DF}_x}\right) \tag{2}$$

where N denotes the total number of documents, and DF_x denotes the number of documents containing the term x.

TF-IDF score is calculated by Eq. (3):

$$\text{TF-IDF} = \text{TF} * \text{IDF} \tag{3}$$

N-grams: The n-grams of textual data are sequence of words or tokens occurring adjacent to each other. Uni-grams and bi-grams are considered in vectorizing the words. Consider an example tweet: *Argentina won the world cup*

This tweet contains uni-grams—"*Argentina*", "*won*", "*the*", "*world*", "*cup*" and bi-grams—("*Argentina*", "*won*"), ("*won*", "*the*"), ("*the*", "*world*"), ("*world*", "*cup*"). Tri-grams, quad-grams, and bigger grams have not been considered as the previous tokens and their significance will get diminished.

After vectorization using Eq. (3), the dataset is divided into two sets, namely the training set and the testing set into an 80:20 ratio. Several machine learning algorithms are used on the training dataset in the initial phase. A logistic regression algorithm is applied to our preprocessed training dataset. A grid search has been done to get the best setting for hyperparameters. The values of C are tuned and initialized class weight as balanced so that, no biasing occurs due to the unbalanced category sizes.

A random forest classifier is used on our training dataset and tuned some hyperparameters like $n_{estimators}(number of estimators)$ to 500, criterion to "*entropy*", and class weight to "*balanced*".

Support vector machine(SVM) has been used to train our training set. The class weight parameter is also tuned and set it to "*balanced*".

Bernoulli Naive Bayes(BernoulliNB) has also been used subsequently. The algorithm was applied on the training set and tuned the hyperparameter alpha(α) and set its value to $1e-6$. Moreover, decision tree classifier and support vector machine with RBF kernel were used and trained with the training data.

4 Results and Analysis

We evaluate our approach by implementing it on multiple datasets and used the metrics, precision, recall, and F1-score. Formulas for precision have been depicted in Eq. (4), recall is shown in Eq. (5), and F1-score is depicted in Eq. (6):

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

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

$$\text{F1-score} = \frac{2 * (\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{6}$$

where TP denotes true positive, FN denotes false negative, and FP denotes false positive.

4.1 Analysis on Benchmark Dataset

We have conducted evaluation of our approach on the benchmark dataset SemEval2016 [12] for stance detection. This benchmark dataset contains records based on different issues raised on Twitter. Tweets are categorized based on *Atheism, Climate Change, Feminist Movement, Abortion Legalization, Hillary Clinton* and *Donald Trump*. The precision, recall, accuracy, and F1-score for the three stances have been analyzed. As depicted in Table 1, our approach is providing great *precision* results for the stances in *FAVOR* and *AGAINST* for the topic of *climate change*. The *precision* results for the topic of *Atheism* are good, and for *Hillary Clinton*, *Feminist Movement* and *Abortion Legalization* results are quite satisfactory. The *recall* results

Table 1 Precision, recall, F1-score, and accuracy on benchmark dataset

Topic	Stances	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Climate change	AGAINST	100	33	56	77
	NONE	68	79	75	
	FAVOR	79	71	80	
Atheism	AGAINST	71	93	78	74
	NONE	67	17	47	
	FAVOR	83	56	79	
Hillary Clinton	AGAINST	73	76	75	72
	NONE	67	65	62	
	FAVOR	68	48	72	
Feminist Movement	AGAINST	60	70	64	57
	NONE	18	8	11	
	FAVOR	45	48	47	
Abortion Legalization	AGAINST	66	76	71	64
	NONE	58	33	42	
	FAVOR	51	59	55	

Table 2 Comparison with the previous method

Approach	F_{avg}
Naive Bayes+ID3-(NER+TF-grams comb) [14]	68.6
SVM-uni-grams [12]	63.31
Our approach	69.33

obtained are also satisfactory for all three stances across all five topics. As depicted in Table 1, it is observed that the *F1-Score* is best for tweets having stance in *FAVOR*, while for tweets, having stance *AGAINST* is moderate in case of *climate change*. For *Atheism* and *Hillary Clinton*, tweets having stance *AGAINST* have comparatively better *F1-score*. The best *accuracy* of 77% is obtained for the topic of *climate change*, while values of *Hillary Clinton* and *Atheism* are also adequate. The lowest *accuracy* of 57% is observed in the case of *Feminist Movement* topic. Our work is compared with the previous method [14] with respect to the SemEval2016 dataset. As depicted in Table 2, it is observed that in comparison with the previous methods, our approach is providing good value for average F1-score(F_{avg}), and the results are satisfactory.

Table 3 Accuracy and details of all classifiers

Classifiers	Hyperparameters	Accuracy (%)
Logistic regression	C = 10, class_weight = 'balanced'	81
Random forest classifier	n_estimators = 500, criterion = 'entropy' , class_weight = 'balanced'	77
Support vector machine (linear)	C = 1, class_weight = 'balanced'	83
Bernoulli Naïve Bayes	alpha = 1e−6	74
Decision tree classifier	max_depth = 128, class_weight='balanced'	79
Support vector machine	Kernel = 'rbf', C = 10	79

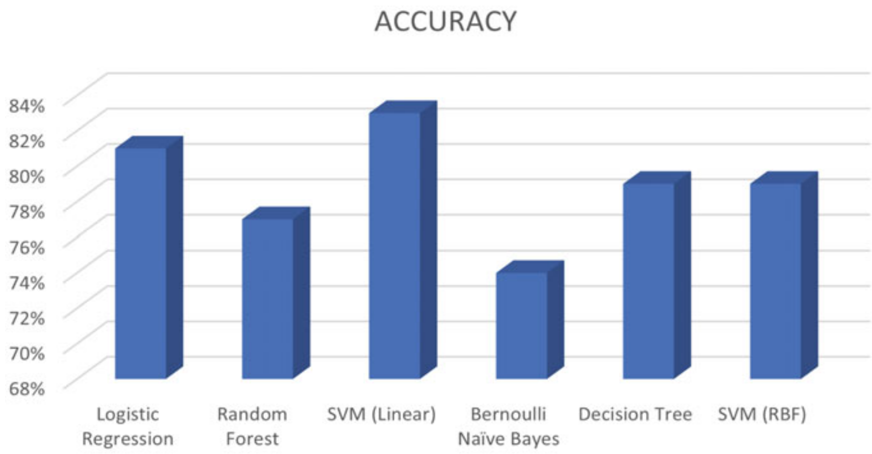


Fig. 2 Accuracy comparison of different classifiers

4.2 Analysis on Collected Dataset

We analyzed our approach for the FIFA World Cup 2022 tweets. After applying the logistic regression model, an accuracy of about 81% is obtained on the test set. In the case of random forest classifier, after training the model, the accuracy value lowers, and an accuracy of about 77% is obtained on the test set. In case of both decision tree classifier and SVM using RBF kernel, an accuracy of 79% is obtained on the test set which is little higher than random forest. In case of SVM, an accuracy of about 83% is estimated on our World Cup test set, which is the best accuracy obtained. For BernoulliNB, an accuracy of about 74% is evaluated on the test set which is comparatively lower than all the other models. The details about the classifiers and the hyperparameters used are summarized in Table 3. The accuracy of the classifiers has been depicted in Fig. 2.

Table 4 F1-score for all classifiers

Classifiers	F1-score
Logistic regression	FAVOR -> 89%, NONE -> 72%, AGAINST -> 78%
Random forest classifier	FAVOR -> 87%, NONE -> 71%, AGAINST -> 68%
Support vector machine (linear)	**FAVOR -> 90%**, **NONE -> 73%**, **AGAINST -> 81%**
Bernoulli Naïve Bayes	FAVOR -> 83%, NONE -> 66%, AGAINST -> 67%
Decision tree classifier	FAVOR -> 86%, NONE -> 69%, AGAINST -> 77%
Support vector machine	FAVOR -> 87%, NONE -> 70%, AGAINST -> 73%

Table 5 Precision and recall values for all classifiers

Classifiers	Precision	Recall
Logistic regression	**FAVOR -> 91%** NONE -> 67% AGAINST -> 84%	FAVOR -> 88% NONE -> 78% AGAINST -> 72%
Random forest classifier	FAVOR -> 90% NONE -> 59% **AGAINST -> 92%**	FAVOR -> 84% **NONE -> 87%** AGAINST -> 54%
Support vector machine (linear)	FAVOR -> 89% **NONE -> 71%** AGAINST -> 86%	**FAVOR -> 90%** NONE -> 76% AGAINST -> 77%
Bernoulli Naïve Bayes	FAVOR -> 81% NONE -> 63% AGAINST -> 73%	FAVOR -> 84% NONE -> 68% **AGAINST -> 84%**
Decision tree classifier	FAVOR -> 86% NONE -> 63% AGAINST -> 90%	FAVOR -> 85% NONE -> 78% AGAINST -> 68%
Support vector machine	FAVOR -> 84% NONE -> 65% AGAINST -> 89%	FAVOR -> 89% NONE -> 76% AGAINST -> 62%

Considering the stance FAVOR, the best value of F1-score at 90% is obtained in the case of SVM(linear). Although logistic regression at 89% is quite close to SVM, the other classifiers linger behind. Bernoulli Naive Bayes performs comparatively worst with an F1-score of 83%.

For stance AGAINST, almost the same trend is observed with the best F1-score of 81% for SVM and the worst F1-score of 67% for Bernoulli Naive Bayes. The detailed F1-score for all the classifiers and respective three stances is depicted in Table 4.

When we evaluate our approach based on precision, logistic regression gives the best result of 91% for stance in FAVOR. But for recall, SVM again gives the highest value of 90%. But for stance AGAINST, the best precision value of 92% is given by random forest classifier and the highest recall value of 84% by Bernoulli Naive Bayes. The precision and recall values of different classifiers are portrayed in Table 5, and the highest values of both the metrics for all three stances are depicted in bold font. The comparison of the precision and recall values of different classifiers have been depicted in Figs. 3 and 4. The confusion matrices for the topics of climate change (Fig. 5a), Atheism (Fig. 5b), Hillary Clinton(Fig. 5c), Abortion Legalization

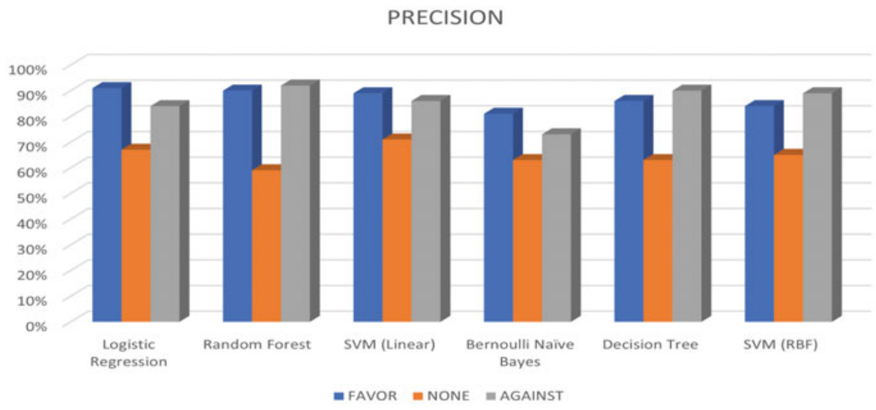


Fig. 3 Precision comparison of different classifiers

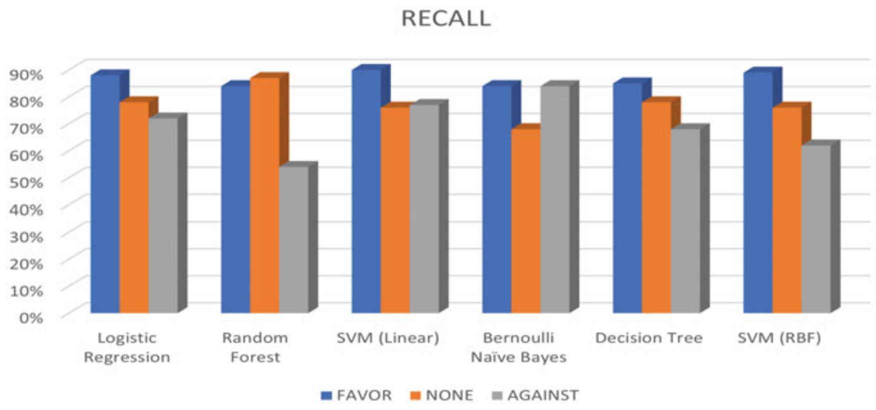
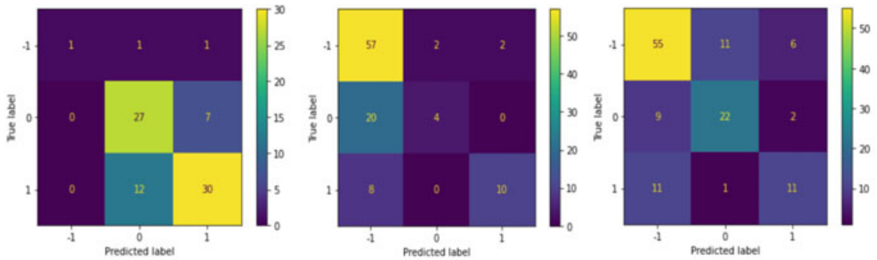


Fig. 4 Comparison of recall values of different classifiers

(Fig. 6a), and Feminist Movement (Fig. 6b) in the benchmark dataset are depicted in Figs. 5 and 6. These confusion matrices are obtained by implementing the SVM approach.

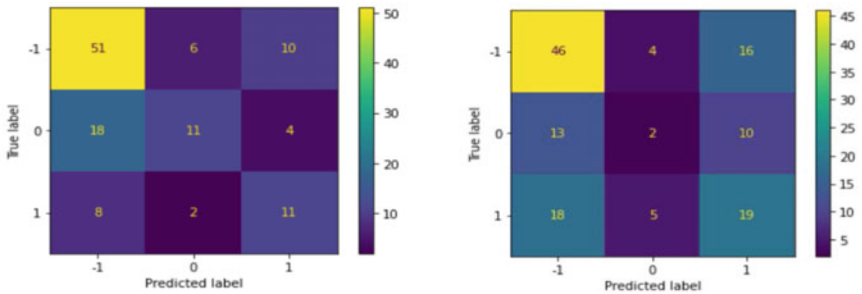
5 Conclusion

Stance classification is the major research topic discussed in this paper. We have collected our FIFA World Cup 2022 dataset from Twitter and preprocessed the raw data using natural language processing techniques for better handling of the information. Different machine learning classifiers like SVM, logistic regression, random forest , Bernoulli Naive Bayes, and decision tree classifies to detect the stance of the tweets



(a) Climate Change (b) Atheism (c) Hillary

Fig. 5 Confusion matrices using SVM approach



(a) Abortion Legalization (b) Feminist Movement

Fig. 6 Confusion matrices on benchmark dataset using SVM

have been used. We compared the accuracy, F1-score, precision, and recall for all the classifiers and observed that SVM performs the finest among all the models. Our approach has been evaluated on benchmark dataset and obtained good results.

References

1. Buongiovanni C, Candusso R, Cerretini G, Febbe D, Morini V, Rossetti G (2023) Will you take the knee? italian twitter echo chambers' genesis during euro 2020. In: Complex networks and their applications XI: proceedings of the eleventh international conference on complex networks and their applications: complex networks 2022-Volume 1. Springer, pp 29–40
2. Burgers C, Beukeboom CJ, Smith PA, van Biemen T (2023) How live twitter commentaries by professional sports clubs can reveal intergroup dynamics. Comput Human Behav 139:107528
3. Cao R, Luo X, Xi Y, Qiao Y (2022) Stance detection for online public opinion awareness: an overview. Int J Intell Syst. https://doi.org/10.1002/int.23071
4. Das S, Chakraborty S (2022) Perception of united nations climate change conference in social networks. In: 2022 IEEE 19th India council international conference (INDICON). IEEE, pp 1–6
5. Dey K, Shrivastava R, Kaushik S (2017) Twitter stance detection-a subjectivity and sentiment polarity inspired two-phase approach. In: 2017 IEEE international conference on data mining workshops (ICDMW). IEEE, pp 365–372

6. Grimminger L, Klinger R (2021) Hate towards the political opponent: a twitter corpus study of the 2020 us elections on the basis of offensive speech and stance detection. ArXiv preprint arXiv:2103.01664
7. Haouari F, Elsayed T (2023) Detecting stance of authorities towards rumors in Arabic tweets: a preliminary study. ArXiv preprint arXiv:2301.05863
8. Hegde SU, Zaiba A, Nagaraju Y et al (2021) Hybrid cnn-lstm model with glove word vector for sentiment analysis on football specific tweets. In: 2021 international conference on advances in electrical, computing, communication and sustainable technologies (ICAECT). IEEE, pp 1–8
9. Hidayatullah AF, Pembrani EC, Kurniawan W, Akbar G, Pranata R (2018) Twitter topic modeling on football news. In: 2018 3rd international conference on computer and communication systems (ICCCS). IEEE, pp 467–471
10. Hou Y, van der Putten P, Verberne S (2022) The covmis-stance dataset: stance detection on twitter for covid-19 misinformation. ArXiv preprint arXiv:2204.02000
11. Kim Y, Billings AC (2017) A hostile sports media? perceived nationalism bias in online sports coverage. Electron News 11(4):195–210
12. Mohammad S, Kiritchenko S, Sobhani P, Zhu X, Cherry C (2016) Semeval-2016 task 6: detecting stance in tweets. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), pp 31–41
13. Price J, Farrington N, Hall L (2013) Changing the game? the impact of twitter on relationships between football clubs, supporters and the sports media. Soccer and Soc 14(4):446–461
14. Tun YM, Myint PH (2019) A two-phase approach for stance classification in twitter using name entity recognition and term frequency feature. In: 2019 IEEE/ACIS 18th international conference on computer and information science (ICIS). IEEE, pp 77–81