

# Comprehensive Analysis of Classifier to Identify Sentiment in Football Specific Tweets

Venkatesh, Keerthi C V, Nagaraj Y, Swetha S

Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, K R Circle, Bengaluru-56001.  
venkateshm.uvce@bub.ernet.in

**Abstract.**The football fans' feelings get unfold during the different phases of the football match and they express their emotions, stance on Twitter. This research work focuses on identifying the stance expressed by the fans on Twitter. The changes in fans' opinions are reflected in a series of tweets written by fans. In this paper, various classification algorithms are used to analyze and categorize sentiments present in tweets posted during the 2018 FIFA World Cup. In this work, a football-specific dataset is created and labeled manually. From the dataset, a lexicon related to football-specific sentiment is created. We use domain-specific lexicons, the TF-IDF feature selection method, Count Vectorizer, and various sentiment classifiers to identify the sentiment expressed by football fans on Twitter. In this paper, the performance of different classifier algorithms is analyzed while determining the hidden sentiment. Our experiment results demonstrate that the Random Forest algorithm exhibits consistent and robust performance compared to other classifiers

**Keywords:**Stance detection, classifier, Twitter, Feature extraction, Lexicon.

## 1 Introduction

The stance detection methods are used extensively in business, stock markets to interpret the opinion of customers towards the product, stock, service, and trends of the market. The sentiment analysis is the most common tool to find out the underlying sentiment present in the message, text, or feedback. The sentiment may be expressed by a layman or domain expert. The sentiment may be written in natural language in a structured or unstructured format. It may be semantically and syntactically incorrect but there are customers' opinions hidden in it. The stance may be in the form of the voice of the customer reviews, survey responses, online submission, and tweets on Twitter. It can be applied in reviewing the emotions expressed by the fans for the entertainment industry and sports. The fans of various sports generate an enormous amount of data. The utility of social media such as Twitter has become a popular platform to express, discuss, and exchange opinions about the sports events that occur. Football events such as the FIFA World cup attract millions of fans around the globe. They express their emotions at different phases of an event such as during

goal-scoring, injuries of players, and also predict the outcome of the match. They use social media as a representative tool to express their emotions through tweets, posts, etc. The sentiment analysis helps in summarizing the intensified emotions exhibited by the fans for the match's win or loss. The stance analysis gives us new insights about how the emotions of fans change at different stages of the sport. Such insights about a certain sports event help in warning the authorities about the possibility of the violence that may occur due to the outcome of it or they might help in analyzing the fandom of a certain player or a team in a particular area. The objective of this paper is to analyze the performance of various classifier algorithms in identifying the hidden sentiments in football-specific tweets of the 2018 FIFA world cup on Twitter social media.

Contributions:

- Prepared our dataset that consists of football-specific tweets which are 2018 FIFA World Cup tweets. All tweets in the dataset preprocessed and labeled manually are used extensively to perform analysis of stance. We collected tweets specific to football events i.e. FIFA World Cup 2018. FIFA dataset consists of about 530k tweets out of which 10,000 tweets are manually labeled.
- Various classifier algorithms are considered to extract hidden sentiment present in football-specific tweets. A comprehensive experiment is conducted to compare various learning algorithms based on various parameters and utilizing a variety of features.
- All football-specific tweets are labeled manually. Count Vectorizer and TF-IDF methods are used to represent textual data in numeric form. Every tweet is tagged with a POS tagger using the tagger tool.
- Applied different classification algorithms on football-specific sentiment dataset and find the classifier that can effectively recognize sentiment in football-related tweets

## 2 Related Work

We are referring to some of the related works that have proposed useful methods to identify sentiments in football specific tweets and performed evaluation on various parameters.

To identify polarity in tweets of the 2014 FIFA World Cup, Barnaghi et al. [1] collected football-specific tweets and labeled them manually. A well-known text classifier algorithm, Logistic Regression is used to find stance or opinion hidden in tweets specific to football events. The unique words of tweets, bigram, and unigram features are included in the dictionary. The proposed classifier achieves less accuracy when it uses a bi-gram feature than that of the uni-gram feature. Sandeep et al. [2] used the same classifier algorithm and similar feature extraction methods. They have used the TF-IDF method and tri-gram features additionally and proved that Logistic Regression has better accuracy when TF-IDF is used to eliminate stop- words. Similarly, Barnaghi et al. [3][18][21] determine the neutrality or individuality present in tweets.

It used an external dictionary and includes other features such as Bigram and Unigram to enhance the performance of the Bayesian logistic regression algorithm and TF-IDF to eliminate the undesirable features from the dataset. The proposed model established a relation between events and sentiments.

Fatimah et al. [4][19] presented a stance detection method that considers a non-domain specific and unbalanced dataset. The proposed method's performance is compared with other well-known classifiers; the proposed classification algorithm classifies data into three categories. Alves et al. [5] considered the 2013 FIFA tweets posted in the Portuguese language and analyzed the tagging of POS for Portuguese words. For manually labeled tweets, SVM and Naive-Bayes learning algorithms are applied to identify the polarity. Performance analysis of SVM, Naive-Bayes classifier concluded that SVM is better than the Naive-Bayes classifier. Similarly, Prabha et al. [6] used a Naive Bayes classifier to prove its efficiency. But Ren et al. [7] used UCI datasets consisting of uncertain data to demonstrate that Naive Bayes has less accuracy than a formula based approach. Kernel density estimation method is proposed to handle uncertain data. Pang et al. [8] prove that Naive Bayes tends to do the worst and SVM tends to the best performance. Xing et al. [9] have modified k-NN classifier that can handle a large sample size of medical data and maintain classification efficiency. The improved k-NN performs clustered based pre-processing and reduce samples by clipping data into a different category. The results reflect that the improved k-NN classifier does better classification for a large sample size. To classify defaults during software development or after the product release and is deployed at the customer, authors in [10] [20] proposed a method that combines k-NN, Linear Discriminant Decision Tree which handles imbalanced data and classify software defaults.

Aloufi et al. [11] have considered tweets related to English Premier League Soccer 2016, and UEFA Champions League Soccer 2016/17. Bag of Words (BoW), lexicon, and linguistic features are used to train the model. Authors have achieved better accuracy of classification when dictionary, Bag-of-Words (BoW), and language-specific features have used the train and test model. Hegde et al. [12] and Hitesh et al. [13] used the Random Forest classifier in performing sentiment analysis which has given better results. Moh et al. [14] used supervised and unsupervised classification methods to analyze the stance of the user. [22] Authors collected 1,50,000 customer stance on movie expressed on Twitter and identified sentiment polarity on a movie using various classifier/learning algorithms. Among all classifiers, the SGD classifier with multi-tier architecture performs better with an accuracy of 87.23%. Aloufi et al. [15] considered various classifiers to categorize and analyze the stance of football fans on the 2016/17 Champions League and 2014 FIFA World Cup dataset. Authors created a lexicon consisting of 3,479 unique words and labeled it according to the stance of fans, and used 54,526 manually labeled tweets to determine the stance of football fans. The experiment results illustrate that SVM exhibited robust and consistent performance compared to RF and MNB classifier.

### 3 Problem Statements

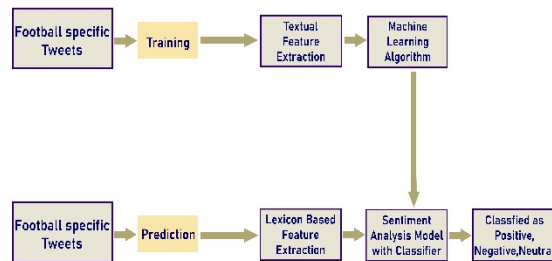
Perform sentiment analysis to find and classify the sentiment of the 2018 FIFA World Cup tweets. Analyze the performance of various algorithms that determine sentiments on sentence level of 2018 FIFA World Cup tweets. The performance analysis of various classifiers concerning sensitivity, accuracy, specificity, precision, recall, and F-score for 2018 FIFA World Cup tweets.

### 4 Proposed Solution

The stance detection algorithms are also known as sentiment analysis models. The domain-specific stance detection algorithms classify the dataset into a favorable, unfavorable, and uncommitted stance. It is widely applied in determining the emotions present in the reviews of the product, marketing, and social media. The sentiment identification can be performed using Natural Language Processing (NLP), an approach that identifies opinions from the spoken and written language.

A stance detection task is modeled as a text categorization problem where data preprocessing, feature selection and classifier algorithm are components of the model. The feature extraction draws out the step that converts raw football-specific tweets to representation form (i.e., feature vectors). These feature vectors are fed to a classifier algorithm to classify the sentiment of tweets, e.g. positive, negative, or neutral. Fig. 1 shows how a machine learning classifier is implemented in this paper.

Fig. 2. represents the activity diagram to exhibit the flow of an operation of the system. The control flows from one operation to another.



**Fig.1.**The architecture of the classifier Model

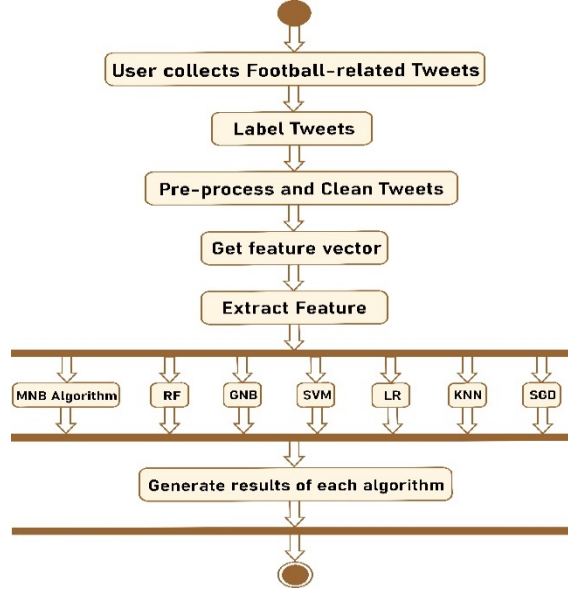


Fig.2. The flow of operation in the proposed system

#### 4.1 Feature Extraction

The classifier algorithm categorizes text into different classes. Since various classifier algorithms require the numerical representation of tweet's text to categorize, therefore, we convert text into a numerical representation. To perform pre-processing, numerical representation, and lexicon, there are various methods: BoW, TF-IDF, and Word2Vec.

##### A. Bag-of-Words (BoW)

The Bag of Words creates a lexicon of the unique words present in a dataset. For each tweet, a set of features is extracted from the lexicon. For every word, the same weight is assigned in BoW. TF-IDF is another approach for constructing a feature vector that assigns a weight for each word. The Term Frequency is determined for a word as shown in Eq. 1 below.

$$TF - IDF_{w_i} = TF_{w_i} \times \log \frac{|DF|}{DF_{w_i}} \quad (1)$$

$$TF - IDFscore = TF * IDF$$

POS features are used to determine whether a word is used as a verb, noun, pronoun, adjective, etc. Every word in a tweet is labeled using TF-IDF and tagged part-of-speech using tagger toolkit is available. POS tag is used in constructing a feature vector.

## 4.2 Lexicon for Football Specific Tweets

The words of Tweets are encoded as integers, or floating-point values, and these values are used as inputs in machine learning algorithms. The stop words are the commonly used words, are discarded before or after processing of raw data. For football-specific tweets, we construct a football-specific lexicon by removing stop words, URLs, hashtags, punctuation symbols, and numerical values from tweets. After this step, the club name, event name that is associated with football, and neutral tweets are discarded. On completion of preprocessing, the vocabulary of unique words and the frequency of each word in tweets is constructed. From vocabulary, the sentiment score of a word in a tweet is determined. *Inverse Document Frequency* (IDF) is the scoring of rare words across the document, is determined and a *score of a word* ( $S_{w_i}$ ) is calculated as in Eq. 2.

$$S_{w_i} = p_{w_i} - n_{w_i} \times \log \frac{n}{df(w_i)} \quad (2)$$

where  $p_{w_i}$  is the probability of a word in a tweet,  $n_{w_i}$  is the number of tweets containing a word,  $n$  is the total number of tweets,  $df(w_i)$  is the number of documents containing a word, and  $S_{w_i}$  is the score of a word. The score of a word is positive for positive tweets, and the score is negative for negative tweets, and the score is zero for neutral tweets.

## 4.3 Dataset

In this paper, the tweets of FIFA World Cup 2018 are collected from Twitter [16]. Out of 530k tweets, we considered 10,000 football specific tweets of the 2018 FIFA world cup. A dataset of 10,000 football specific tweets are manually labeled and split into training and testing data set in a ratio of 60:40.

## 4.4 Classifier Algorithm

In this section, we discuss various classification algorithms that include Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM), Random Forest (RF), Gaussian Naïve Bayes, Logistic Regression, k-NN (k- Nearest Neighbor), and Stochastic Gradient Descent (SGD).

### A. Multinomial Naïve Bayes (MNB)

In a Naive Bayes classifier, all features are independent; the occurrence of a feature is independent of another feature for a given class of variables. For the multinomial event model, the event frequencies are generated by a multinomial ( $p_1, \dots$ ) where  $p_i$  is the probability of occurrence of an event, similar to the occurrence of a word in a document. The feature vector is given as  $\mathbf{x}_i = (x_1, x_2, \dots, x_n)$ , the probability of a word or event  $x_i$  in a document is given by Eq. 3.

$$p(\mathbf{x} | C_k) = \frac{(\sum_i x_i)!}{\prod_i x_i!} \prod_i p_{k_i}^{x_i} \quad (3)$$

In log-space, Multinomial Naïve Bayes can be represented as a linear classifier. The event frequency is zero if the event does not occur in a document. It is given as in Eq. 4.

$$\begin{aligned} \log p(C_k | x) &\propto \log \left( p(C_k) \prod_{i=1}^n p_{k_i}^{x_i} \right) \\ &= \log p(C_k) + \sum_{i=1}^n x_i \cdot \log p_{k_i} \quad (4) \\ &= b + w_k^T x \end{aligned}$$

where  $b = \log p(C_k)$  and  $w_{k_i} = \log p_{k_i}$

### B. Support Vector Machine (SVM)

SVM classifier defines a boundary that is used to categorize generalized dataset into different groups. We are using an SVM classifier to define a boundary that can be used to categorize domain specific dataset. We are using an SVM classifier to easily classify football-specific tweets into the correct category. The optimal decision boundary is referred to as a hyperplane. SVM selects a set of vectors that would be used in defining a hyperplane. SVM classifier chooses a hyperplane that categorizes nearby and interleaved tweets. In this paper, we use the RBF kernel for training model generation. The hyperplane is determined by solving Eq. (5).

$$\arg \min_{w,b} \frac{1}{2} \|w\|^2 \quad (5)$$

Constraint to  $y_i (w \times x_i + b) \geq 1, 1 \in [1, n]$

where  $x_i$  is feature vector,  $y_i \in \{+1, -1\}$

### C. Random Forest Algorithm

Random Forest is a classification algorithm that constructs a decision-tree. It follows a bagging approach to construct a tree in which the decision tree can be expanded to the maximum level. It is an ensemble decision-tree categorization technique; a randomly selected dataset is used in training. A decision-tree is created based on trained instances of bootstrapped and randomly chosen feature sets. The strong decision tree is constructed from a small set of individual learner tree. The prediction is determined by considering votes cast by each learner tree. The prediction is determined by considering votes cast by each tree. Random Forest performs consistent and robust in presence of noise and overcomes the problem of overfitting. RF is widely used due to the capability of handling large datasets and is suitable for multiclass classification.

### D. Gaussian Naïve Bayes

A Gaussian distribution is also called Normal distribution. For continuous data, values associated with each class are distributed according to Gaussian distribution. For example, to assign value to a continuous attribute,  $x$ . First, categorize the dataset

8

by class. Second, compute the mean and variance of  $x$  in each class. It is shown in Eq. (6).



**Logistic Regressions** In Logistic Regression, the target variable  $y$ , can take non-continuous values for a given set of features  $x$ . It is very useful to learn the influence of several independent variables on a single outcome. Similarly, in multinomial Logistic Regression, the output variable has more than two discrete outputs, it uses the softmax with cross-entropy loss function instead of the sigmoid function. The logistic regression is expressed as in Eq. 7.

$$y = \frac{e^{b_0 + b_1 \times x}}{(1 + e^{b_0 + b_1 \times x})} \quad (7)$$

Where  $v$  is expected output.  $b_0$  is a bias term,  $b_1$  is the input variable coefficient.

#### E. K-Nearest Neighbor(k-NN) Algorithm

The k-NN takes test data and training dataset as input from the dataset to choose an appropriate value for data points (i.e. k). Using any method (Manhattan or Euclidean or Hamming distance), distances between the test dataset and every row of the training dataset is calculated. Sort them according to the distance. The output variable is categorized based on neighbors' plurality vote; the output variable is labeled with a class that is common among its k nearest neighbors.

#### F. Stochastic Gradient Descent (SGD)

In SGD, it shuffles the data set and picks one data point from the whole data set at each iteration to the gradient. It is also possible to compute the gradient for a sample of a small number of data points instead of just one point at each step, it is called "mini-batch" gradient descent. To determine the gradient of the objective function for each feature, pick a random initial value for the parameters 'y' for features like  $x_1$ ,  $x_2$ , etc. The gradient function is updated by plugging in the parameter values. Find step sizes for each feature and new parameters.

### 5. Experiment Results and Analysis

The classifier algorithms can identify the sentiment hidden in football-specific tweets of the 2018 FIFA world cup. The collected dataset is used to train various classifier algorithms and compare the performance of each with respect to accuracy, precision, recall, sensitivity, specificity, confusion matrix, and F1 score. A total of 10,000 Football specific tweets of the 2018 FIFA world cup are selected from 530k tweets of the world cup and manually labeled. The dataset is split in the ratio of 60% - 40% for training and testing.

#### A. Performance Parameters

The performance parameters used to evaluate the algorithms are defined as follows.

A confusion matrix contains rows and columns; an instance of the predicted class is rows of matrix and columns represent actual class instance.

$$\text{Accuracy} = (TP+TN) / (TP+FP+FN+TN)$$

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}),$$

Sensitivity= TP / (TP+FN)

Specificity = TN / (TN+FP).

## B. Experiment Results

In this section, the performance of different parameters while classifying football-specific tweets into positive, neutral, or negative class are discussed. Figure 3 illustrates the performance of the Random Forest algorithm. It shows the confusion matrix, classification report, and result of various other performance metrics. As shown in Fig. 3, similar results are obtained for the remaining classifiers.

Fig. 4.demonstrates the performance of various classifier algorithms with respect to the accuracy, macro F1 score, and weighted F1 score. It is observed that the Random Forest algorithm has an accuracy of about 82.25%, 76% macro average F1 score, and 82% weighted average F1 score. Compared to all the other classifier algorithms, Random Forest performs better.

Fig. 5.demonstrates the performance of various classifier algorithms for different evaluation parameters such as macro average Precision, weighted average precision, macro average Recall, and weighted average Recall.

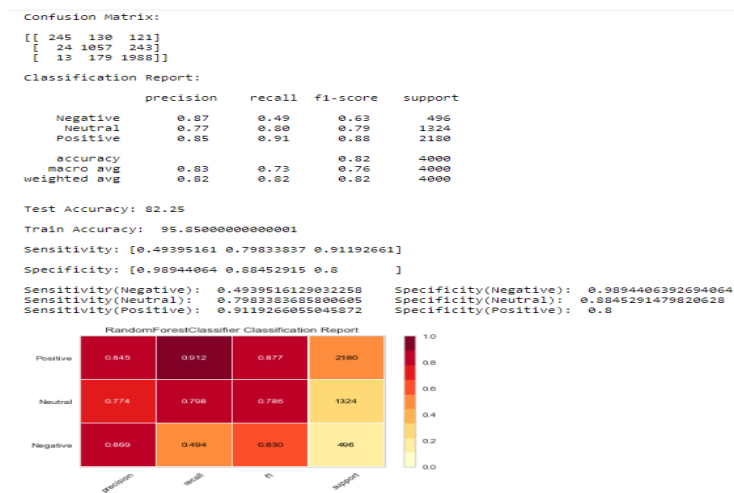


Fig. 3. Confusion Matrix of Random Forest Classifier

Fig. 6 shows the performance of Multinomial Naïve Bayes classifier for 10,000 football-specific tweets which classify them into positive (54.1%), negative (33.8%), and neutral (12.2%) class. Fig. 7 show the confusion matrix of the k-NN algorithm using the CountVectorizer lexicon for football-specific tweets.

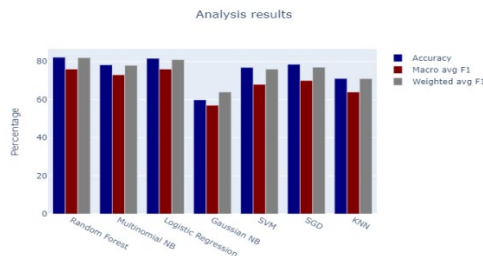


Fig. 4. Performance of classifier algorithms with respect to Accuracy, macro F1 score, Weighted F1 score.

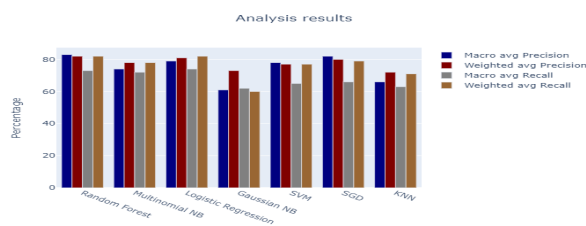


Fig. 5. Performance of classifier algorithms with respect to macro average Precision, Recall, and weighted average Precision, Recall.

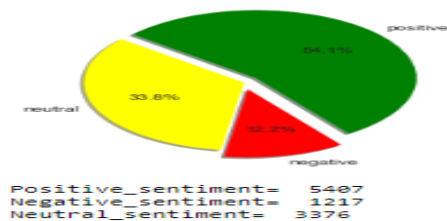


Fig. 6. Performance of MNB classifier

Table 1 illustrates the performance of various classifier algorithms. It is observed that the Random Forest algorithm achieves accuracy of about 82.25% using CountVectorizer lexicons. Similarly, it is observed that other classifiers have achieved more accuracy using the CountVectorizer lexicon than TFIDF lexicons except for SGD which uses TFIDF Transformer.

Table 1. Performance of classifier algorithms for various performance parameters

Classifiers	Feature extractor	Accuracy (In %)	Micro average			Weighted average		
			F1 score	Precision	Recall	F1 score	Precision	Recall
Random Forest	Count Vectorizer	82.25	76	83	73	82	82	82
	TFidf Vectorizer	81.39	76	81	73	81	82	81
Multinomial NB	Count Vectorizer	78.25	73	74	72	78	78	78
	TFidf Vectorizer	76.52	69	76	66	76	76	77
Logistic Regression	Count Vectorizer	81.07	76	79	74	81	81	82
	TFidf Vectorizer	79.85	72	82	69	79	80	80
Gaussian NB	Count Vectorizer	59.87	37	61	62	64	73	60
	TFidf Vectorizer	55.2	53	60	60	60	74	55
SVM	Count Vectorizer	76.87	68	78	63	76	77	77
	TFidf Vectorizer	78.52	70	82	66	77	80	79
KNN	Count Vectorizer	71.07	64	66	63	71	72	71
	TFidf Vectorizer	69.07	63	78	63	69	79	69

## 6. Conclusions

In this work, the performance of various classifier algorithms such as Multinomial NB, Gaussian NB, Random Forest, SVM, Logistic Regression, k-NN, SGD on a football-specific dataset consisting of 10,000 manually labeled tweets from FIFA 2018 is analyzed. The results have shown that RF (Random Forest) with Count Vectorizer feature extraction has achieved stable and robust performance compared to all other learning algorithms. Apart from RF, Logistic Regression (LR) has achieved better performance compared to the remaining algorithms. SGD, Multinomial NB, SVM, k-NN algorithms have shown quite good performance. All algorithms can achieve better performance using CountVectorizer except SGD which uses TF-IDF Transformer.

## References

1. P. Barnaghi, P. Ghaffari, and J. G. Breslin, "Text Analysis and Sentiment Polarity on FIFA World Cup 2014 Tweets", In Proceeding of ACM Conference on SIGKDD, vol. 15, pp. 10-13, 2015.
2. Sandeep Nigam, Ajit Kumar Das, Rakesh Chandra Balabantaray, "Machine Learning-based Approach to Sentiment Analysis" in Proceedings of International Conference on Advances in Computing", Communication Control and Networking ICACCCN2018, pp. 157-161, 2018.
3. Barnaghi, P. Ghaffari, and J. G. Breslin, "Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment", in Proceedings of 2nd IEEE International Conference on Big Data Computing Service and Applications (BigDataService) pp. 52-57, 2016.
4. F. Alzamzami, M. Hoda, and A. E. Saddik, "Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation," In IEEE Access, vol. 8, pp. 101840-101858, 2020.
5. Alves, A. Luiz, C. A. A. Firmino, M. G. de Oliveira, Paiva, "A Comparison of SVM versus Naive-Bayes Techniques for Sentiment Analysis in Tweets: A Case Study with the 2013 FIFA Confederations Cup" in Proceedings of the 20th Brazilian Symposium on Multimedia and the Web. 2014., New York, pp. 123-130, 2014.

6. Surya Prabha PM, Subbulakshmi B, "Sentimental Analysis using Naïve Bayes Classifier", in Proceedings of 2019 International Conference on Vision Towards Emerging trends in Communication and Networks (ViTECoN), pp.1-5, 2019.
7. J. Ren, S.D. Lee, X. Chen, B. Kao, R. Cheng, and D. Cheung, "Naive Bayes Classification of Uncertain Data" in Ninth IEEE International Conference on Data Mining, pp. 944-949, 2009.
8. Bo Pang, Lillian Lee, and ShivakumarVaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques [C]", in Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing, pp. 79-86, 2002.
9. W. Xing and Y. Bei, "Medical Health Big Data Classification Based on KNN Classification Algorithm," In IEEEAccess, vol. 8, pp. 28808-28819, 2020.
10. I. Tumar, Y. Hassouneh, H. Turabieh, and T. Taher, "Enhanced Binary Moth Flame Optimization as a Feature Selection Algorithm to Predict Software Fault Prediction," in IEEE Access, vol. 8, pp. 8041-8055,2020.
11. S. Aloufi, F. Alzamzami, M. Hoda, and A. E. Saddik, "Soccer Fans Sentiment Through the eye of Big Data: The UEFA Champions League as A Case Study", in Proceedings of IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 244-250, 2018.
12. YashaswiniHegde, S.K. Padma, "Sentiment Analysis using Random Forest ensemble for Mobile Product Reviews in Kannada" in Proceedings of IEEE 7th International Conference on Advanced Computing (IACC), pp. 777-782, 2017.
13. MSR Hitesh, VedhosiVaibhav, Y.J AbhishekKalk, SurajHarshaKamtam, SantoshiKumari, "Real-Time Sentiment Analysis of 2019 Election Tweets using Word2vec and Random Forest Model" in Proceedings of 2nd International Conference on Intelligent Communication and Computational Techniques (ICCT), pp.146-151,2019.
14. Melody Moh, AbhitejaGajjala, Siva Charan Reddy Gangireddy, Teng-Sheng Moh, "On Multi-tier Sentiment Analysis Using Supervised Machine Learning" in Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 341-344, 2015.
15. SamahAloufi, Abdulmotaleb El Saddik, "Sentiment Identification in Football-Specific Tweets" in IEEE Access, vol.6, pp. 78609- 78621, 2018.
16. Dataset source- <https://www.kaggle.com>
17. Websites-<https://www.wikipedia.org/>,  
<https://towardsdatascience.com/>,<https://medium.com/>, <https://www.semanticscholar.org/>,  
<https://www.researchgate.net/>
18. S. T. Ahmed and K. K. Patil, "An investigative study on motifs extracted features on real time big-data signals," 2016 International Conference on Emerging Technological Trends (ICETT), Kollam, 2016, pp. 1-4, doi: 10.1109/ICETT.2016.7873721.
19. K. D. Singh and S. T. Ahmed, "Systematic Linear Word String Recognition and Evaluation Technique," 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 2020, pp. 0545-0548, doi: 10.1109/ICCSP48568.2020.9182044
20. Kumar, S.S., Ahmed, S.T., Vigneshwaran, P. et al. "Two phase cluster validation approach towards measuring cluster quality in unstructured and structured numerical datasets". J Ambient Intell Human Comput (2020). <https://doi.org/10.1007/s12652-020-02487-w>
21. M. Anathi, K. Vijayakumar , "An intelligent approach for dynamic network traffic restriction using MAC address verification", Computer Communications,Elesvier,5 February 2020.

22. K. Pradeep Mohan Kumar, M. Saravanan, M. Thenmozhi ,K. Vijayakumar, “ Intrusion detection system based on GA-fuzzy classifier for detecting malicious attacks”, wiley, Feb 2019.