

SENTIMENT ANALYSIS OF TWITTER TWEETS TO IDENTIFY THE NEGATIVITY FACTORS

Bilal Ahmed Chandio¹, Muhammad Aimal¹, Maheen Bakhtyar¹, Junaid Babar¹, Maheen Afzal²,

¹Computer Science & IT department, University of Baluchistan, Quetta.

²Sardar Bahadur Khan Women University, Quetta.

*Correspondence: bilal.csit@um.uob.edu.pk

ARTICLE INFORMATION

Citation: Bilal Ahmed Chandio, Muhammad Aimal, Maheen Bakhtyar, Junaid Babar, & Maheen Afzal. (2022). SENTIMENT ANALYSIS OF TWITTER TWEETS TO IDENTIFY THE NEGATIVITY FACTORS.

Pakistan Journal of Emerging Science and Technologies, 2(2), 11.

<https://doi.org/10.5281/zenodo.5814103>

Received: 21st Nov-2021

Revised and Accepted:

26-December-2021

Published On-Line

03-January-2022

***Corresponding Author:**

Bilal Ahmed Chandio:

bilal.csit@um.uob.edu.pk

Original Research Article

ABSTRACT

An ample amount of data is being daily generated on Twitter where people express their sentiment and emotions. Analyzing those sentiments using NLP (Natural Language Processing) techniques would help the individuals and organizations to scientifically process the prevailing sentiment. The growing negativity on social media has compelled the researchers to innovate a scientific method to analyze the user sentiment. This paper is using approaches of sentiment analysis by applying different methods and algorithms. Our proposed method is based on logistic regression to classify the intensities towards the emotions extracted from tweets. The dataset collected from twitter is narrowed down to twitter Pakistan tweets and then trained our model on the training dataset and later tested the model on testing dataset where different accuracies are experimented by logistic regression algorithm.

Keywords: Classification, Sentiment Analysis, Logistic Regression, BOW (bag of words), NLP, intensities, emotions



Pakistan Journal Emerging Sciences and Technologies (PJEST) by [Govt. Islamia College Civil Lines Lahore, Pakistan](#) is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).

Introduction:

In today's world, everyone is interested in someone's Emotions about the facts and products they will buy through online platforms. Emotion is deliberated as 'A strong feeling deriving from one's circumstances, mood, or relationships with others. American Psychological Association has quoted 'emotion' as "A complex pattern of changes, including physiological arousal, feelings, cognitive processes, and behavioral reactions, made in response to a situation professed to be personally important". As a whole, 'emotion' is the feeling or reaction that people

have on a certain event. 'Happy', 'Sad', 'Angry', 'Fear' are few instances of emotions which someone could express against any issue or product. 'Emotions' and 'sentiments' these two terms are perceived as an interchangeable term however, 'sentiment' represents a more general idea which leads to polarity which is either positive, negative or neutral.

Emotions are a complex psychological state that involves different components that have several subjective experiences, stated physiological response and counter behavior that can be expressive or maybe not. Individual emotional states, often known as moods, are important in the expression of thoughts, ideas, and views, which influences attitudes and conduct [1]. Emotion is a complex, multifaceted feature that represents a person's personality and behavioural characteristics [2]. People express their feelings about a variety of situations, events, people, the environment, and even the smallest things in their everyday lives. The emotions are compiled to be measured because they contain those human characteristics that depict actual psychological conditions. The human psychological reactions can be positive or negative counting positive reviews about the product can intensify the level of positivity of that product. Counting the negative reviews about that product can intensify the negativity about that product. Different sentiment analysis approaches can be used to measure the intensities of positive and negative reviews. The emotions are expressed on social media by using text and graphical content; people share their thoughts and ideas through social media posts, status, comments, blogs on websites and other mediums. Some basic emotions for the polarity of negativity are fear, anger, and sadness. The intensity of the positive and negative reviews is measured using natural language processing methods.

The social media platforms have emerged as a stage to furnish an arena for the people to express their opinion, sentiment and perspective on different issues subjects. People share their thoughts, mental state, moments, stand on specific social, national, international issues through text, photos, audio and video messages and posts. Despite the availability of different communication methods, the text is still a prominent form of communication [3]. An estimated 61.34 million people in Pakistan are internet users and connected to various social media platforms. In Pakistan people tweets almost on daily bases and consequently generates an ample amount of text data. This overwhelming increase of data has posed a serious challenge for researchers to investigate an efficient method for analyzing such text data.

Analyzing and detection of emotion from a piece of text has been major research in NLP during the last decade. A variety of techniques have been proposed by the researchers to mitigate this research problem. To correctly identify the emotion, it is necessary to understand the complexity of human emotion and the challenges. A major challenge while extracting the emotion is to mitigate the piece of text having multiple emotions. When emotion is so implicit in a text, automatic emotion detection is practically impossible [4]. Many sarcastic sentences are difficult to discern even for other people, let alone for a machine to detect accurately.

We propose logistic regression (which is a supervised machine learning algorithm) for detection of emotion. Logistic regression which is usually used for binary classification problem aimed at establishing the relationship between variables so as to efficiently determine the class. In contrast to linear regression, logistic regression can provide a nonlinear border between two categories. If the two classes in question could be divided linearly, linear regression might be

employed successfully for facial recognition. The logistic regression having capability of non-linear boundary enables us classify the data which is even not linearly separable. We have employed the logistic regression for multiclass problem of emotion detection. In our proposed scheme the data is initially pre-processed by cleaning the unnecessary symbols and punctuations and later tweets are tokenized and stemmed. Subsequently, the processed tweets are placed in a vector space of bag-of-words model and are ready to detect the emotions using logistic regression. The proposed methodology is also illustrated in Fig. 1 whereby, it can be observed that soon after classification the resultant probabilities are evaluated tested.

Limitation of the Study:

This study shows that there is no limitation imposed on social media regarding age. The social media platforms allow to create an account only if the person age is 18 or over 18. It will not allow to create an account on social media if the age is under 18.

In our study, all are equally male and the female difference is not treated as gender difference we only need the comments based on emotions for research purpose. In today's world the attraction of social media platforms like Twitter, is a social media platform designed to know the interest of persons/people to according to any topic or online marketing where a person can tweet against a topic according to his/her own assumptions. From these tweets, we can explore a product, or any topic according to their conversations, By collecting these tweets we will be able to know the feedback of that product in online marketing like that we can collect the huge amount of tweets/comments of political parties, or any political persons from any social media using different techniques, With the help of these tweets we can adopt a methodology to intensify the level of feedback according to that political parties or any political persons. Exploring someone's personal emotions to know the personal interest of that product or any political terms, through which they can know that via a person is happy, sad or fear and angry, if a person is angry then they will be in interest to know the level of that anger.

The researchers use Twitter and other social media platforms for analysis and calculating emotions and sentiments. The tweets on Twitter are shared, favorite and retweeted to endorse or agree on the statement shared. This paper analyses the negative emotion from the Twitter Pakistan dataset and uses the present technique incorporating new methods. The article is organized as follows the section 2 mentions existing work and achievements on sentiment analysis. Section 3 discusses the methodology we designed, and in section 4, the result and discussion is made, and the last section includes conclusion, limitation and future research work.

Related Work

The social media is the only platform where you can know about any person or political terms and products. This study shows that the author wants to explain that with the help of social media we can collect textual data about any topic, person, or any other interesting things which can attract you by its design, we can adopt different techniques and methodologies to intensify the level of that product or etc. Furthermore, the author suggests that by applying various algorithms like

classifications or etc. [5]. Associating one word with the full sentence is another task of NLP emerging in today's knowledgeable tasks, N-Gram is the technique where one word is associating with another word or sometimes with full sentence to measure the degree of full-sentence via its positive or negative [6]. Positive Sentence according to N-Gram means less negative and a negative sentence means less positive, with this evolving technology we can intensify the exact level of emotions by N-gram measuring emotions. [7] [8]. Correlation between sentences and combination of words is getting high results comparing to a single model, Combination of words are known as a thesaurus in NLP different word combination results better comparing to single sentence for text-based algorithms. With the need for thesaurus, we can achieve a model which will results high in classification for use of classification models. CNN known as a convolutional neural network that is totally based on neuron technology can work or achieve better results by applying thesaurus on sentences. [9] [10].

In this technology era Sentiment expression presenting pattern as closing work in data mining. When analyzing sentiment various envision related to any topic can be found. In online platforms where we have distributed or cloud computing. People can share their assets using cloud technology. The structure of web-based technology uses a server for storing personal or any information a user wants. Furthermore, the author decided to explore Twitter cloud information using techniques of data mining. [11]. Analyzing meaningful data from social media unstructured data is undoubtedly a tedious job. On various social networking sites, a vast number of tremendous datasets regarding opinioned data is available, due to this large amount of data opinion mining is today's most hot topic for researchers [12]. One of the most popular sites for microblogging is Twitter, used by literally many peoples. Because of its opinioned data and free of cost, it is merged in sentiment analysis most searched topics. [13].

Social networking enables online learning based on the popularity and powerful station of the internet, where users can share ideas and opinions. Social media has a large amount of text in the type of tweets for sentiment analysis. For the classification of tweets, this paper addresses the sentiment analysis technique to keeping the sentiment positive, negative, and neutral. In the market to know the review of shoppers for any product, it became important to seek the general public emotion about the shoppers, predictions of political elections, socio-economic activities and the stock market, etc. This research defines a sentiment classification method developing a classifier to automatically identify the unknown tweets from Twitter using CNN (convolutional neural networks) [14]. Machine learning is an approach for sentiment analysis on Twitter, solving the problems of NLP machine learning is helpful by combining general algorithms, training a huge amount of data on samples for basic classifier rules [15] [16] Twitter sentiment analysis focuses on analyzing the feelings of tweets and feeding that information to machine learning algorithms to check its accurate results. The result of this research shows that machine learning algorithms like SVM and naïve Bayes have the best accuracy models, for baseline learning algorithms it should be considered. [17].

People sharing their daily data on social media platforms in a form of tweets. The tweet is a text-based expression, but a tremendous amount of insights can be harvest from these unstructured tweets. Some organizations are interested to know the opinions of the public about the services and products they are dealing with, And to know the sentiment of these tweets a method known as sentiment analysis is used which is related to NLP and textual data mining. This paper addresses

the techniques of SA to know the sentiment from tweets. With the invention of smartphones, it is very easy to access social media platforms. The author uses the English language in this research for sentiment analysis approaches. [18]. Prediction about stock markets is a highly demanding task due to its randomness and notable returns in the stock market. Researchers working to improve accuracy from social media and consideration news. This research defines the estimations of whether is terms of sentiment analysis derived from Twitter posts. To enhance the accuracy of classification researcher used n-grams along with two strategies are applied. For the processing of Twitter data Spark has been selected and for analyzing Apache Flume is utilized. With respect to current social trends and opinions, Twitter is considered a rich source of real-time data. This work defines a methodology that forecasts the movements of initial prices of the shares and can be applied to various companies. [19].

As a storm of Coronavirus pandemic taken the world as well as in social media. This paper defines the statistical analysis on Twitter messages of Covid-19, two types of experimental studies performed on, first on sentiments of individual tweets and the second one are on word frequency. To look over word frequency used on the site for characterizing the pattern or trends, the spreading of this disease in social media reflected a psychological threat on Twitter users. Sentimental research, for example in industries, has many uses across different contexts to get input across products from which businesses can collect consumer feedback and social media comments In this technical paper we demonstrate how sentimental analyzes are implemented and how to connect to Twitter and perform sentimental analytical queries. We perform experiments on various questions from politics to mankind and give fascinating results. We have found that the neutral feeling for tweets is dramatically strong, clearly showing the flaws of the current plays. [20]. Transformation of traditional media to social media platforms, for communication and updated companies, switched from lengthier public announcements to Twitter messages (tweets). Many businesses also rely on Twitter to communicate with stakeholders. To help the researchers to better understand the “Viral” tweets, the author performed an exploratory analysis on dataset counted 18000 tweets from multiple areas having content, sentiment as fundamental events, to explain tweets performance author performed sentiment analysis by adopting a model. [21].

This study mainly focuses on sentiment analysis of Twitter data from public transport to measure the opinions on services provided by public transport and to classify these opinions via neutral, negative, and positive opinions. By analyzing the tweets, the researcher describes factors of lacking the use of public transport and also the factors that make the use of public transport and facilities that provides. This research will help public transport to improve their facilities and reduced traffic jams and pollutions by using public transport for transportation. Accuracy is calculated with Naïve Bayes and logistic regression. [22]. Investors and researchers widely apply sentiment analysis in the finance domain, but most research is conducted on English text. In this research, the author represents a framework that analysis and visualizes a dataset consisting of Arabic tweets from the Saudi stock market, and applying machine learning algorithms for sentiment analysis. For prediction and training, Twitter API is used for collecting off-line data and Apache-Kafka was used for real-time streaming tweets. Five different machine learning classifiers with different feature extraction including word embedding (word2vec) and the traditional BoW methods were used for experimental purposes. The highest accuracy for the sentiment

classification of Arabic tweets was 79.08%. This result was achieved with the SVM classifier combined with the TF-IDF feature extraction method. [23].

Different SA classification approaches can solve the problems related to the network geology problems of networking-related tasks. [24]. SA of Twitter tweets related to Arabic linguistics makes it very convenient to solve the tasks or problems of Arabic languages using multiclass approaches of sentiment analysis. [25]. In recent years sentiment analysis using Twitter data by machine learning is a very popular topic. Addressing the problem of sentiment analysis using critical events like natural disasters or social movements. Performing the sentiment analysis on two datasets in Spanish: the 2010 Chilean earthquake and the 2017 Catalan independence referendum, considering the Bayesian network classifier. Controlling the number of edges in the Bayesian network classifier from training examples to reduce the number of edges, implementing the Bayes factor for this purpose getting the most realistic approach for realistic networks. The competitive results show the effectiveness of Bayes factor measures with its predictive results comparing with its support vector machines and random forests, with a specific number of training samples. As a result, networks ensuring to identify the relationship between words, adopting the qualitative interest of historical and social events comprehend the main event of features dynamics. [26]. Many popular machine learning and deep learning approaches in their specific applications, such as Support vector machines (SVM), random forests for classification, hidden Markov models, Bayesian networks, and others, were carried out in genomics, proteomics, structures biology, and several different domains [27]. The Naive Bayes algorithm calculates the probability of data point whether it belongs to a specific category or not [28] [29]. In this type of algorithm, labels are used, and data is analyzed. The main purpose of this algorithm is to get the base value and accuracy of the dataset. The K-Nearest neighbors are influential because they come from real-life and solve regression and classification problems. The KNN is following the data point that has value in the majority. The KNN is working on the importance of its data point [30].

Methodology

The proposed methodology uses the dataset that is collected from Twitter Pakistan. That data is filtered and divided into multi-classes to capture the annotation of the tweets, whether the words are creating a negative or positive sense. The purpose of the methodology is to capture negative connotation tweets in which three fundamental emotions are highlighted: sadness, anger, and fear. Those tweets are filtered, and further, the text preprocessing method is used where the stop words, lemmatization, stemming off the words and tokenization is applied to make the dataset able to be used in the logistic regression model.

The logistic regression model is used for predictive analysis. The classes present in the dataset are analyzed and classified from the corpus used in the method. The model is tested and evaluated on the dataset, and the results are displayed after testing on multiple classes in the dataset. Logistic regression is mathematically defined in Eq. (1).

$$g(\mathbf{J}(\mathbf{x})) = \omega + \epsilon x_1 + \beta x_2 \quad (1)$$

Here, $g()$ is the link function, $J(x)$ is the expectation of target variable and $\hat{\omega} + \hat{\epsilon}x_1 + \hat{\beta}x_2$ is the linear predictor ($\hat{\omega}$, $\hat{\epsilon}$, $\hat{\beta}$ to be predicted).

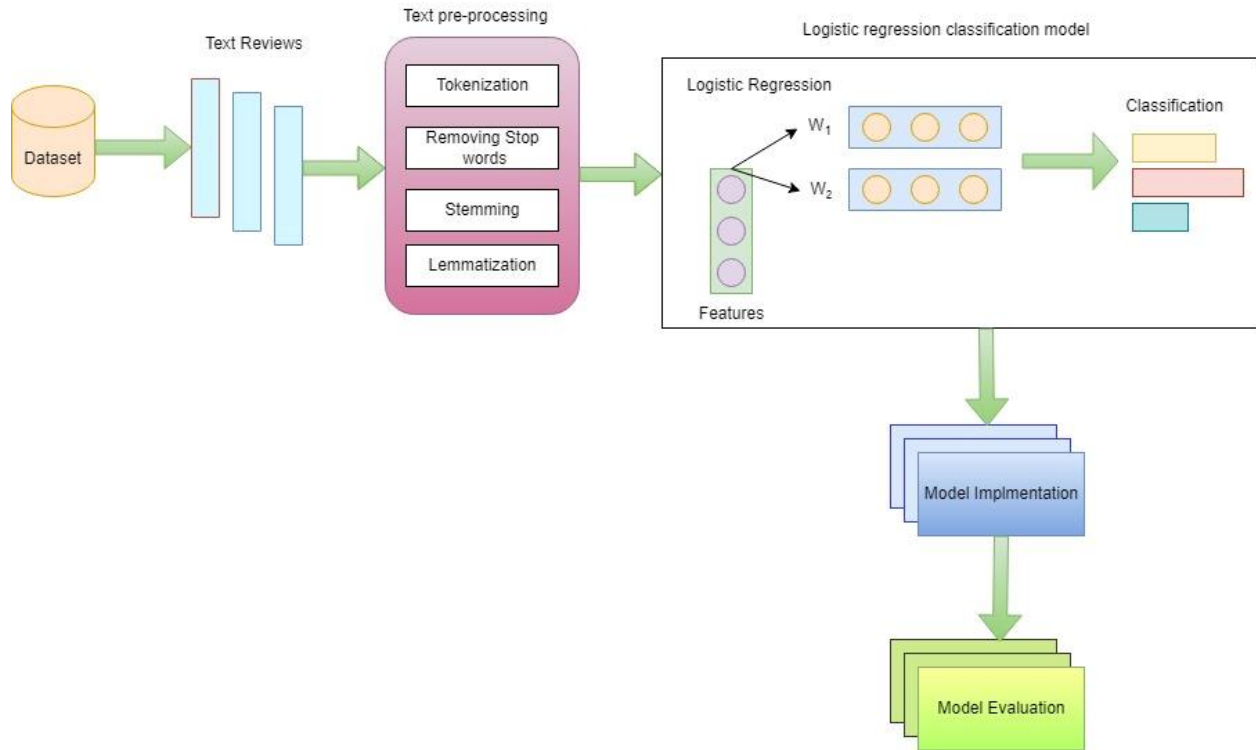


Fig. 1: Methodology to visualize

Results and Discussion

Frameworks:

Using Twitter API (Tweepy) for tweets collecting process and after collection phase now it's the second phase to select the areas of Pakistan for the process of creating a dataset.

Collection:

Tweets are extracted using only three emotions Sadness, Fear, and Anger, ten factors were identified using these emotions. These ten factors are the reasons, these reasons were converted to eight classes to analyze the dataset using multiclass approaches/ techniques. This research showing only three emotions like Fear, Anger, and Sadness.

Classification:

These ten classes are basically extracted from emotions like Sadness, Fear, Anger which are used to collect the data to create a dataset for this research. From Anger emotion when extracting tweets from Twitter classes like Crime, Terrorism, Politics, and Injustice were

categorized in this emotion, from fear emotion classes like Failure, Corruption, Economics, anxiety were categorized in this emotion, a Classes like the social aspect, Failure, Economics, poverty and injustice were categorized from Sadness emotion. These eight classes are basically extracted from these three emotions (Sadness, Anger, and Fear). These ten classes are the reasons or factors identifying the negativity from social media.

TABLE I: Dataset Description

Classes	Training Dataset	Testing Dataset	Class Distribution
Politics	395	131	20.41%
Injustice	297	99	15.37%
Crime	252	84	13.04%
Economic	234	76	12.03%
Failure	209	70	10.83%
Terrorism	202	68	10.48%
Social Aspects	195	65	10.09%
Corruption	150	50	7.76%
Anxiety	297	99	15.37%
Poverty	150	50	7.77%

TABLE II: Logistic Regression Testing accuracy

Classes	Precision	Recall	f1-score	Support
Politics	0.66	0.61	0.63	76
Injustice	0.79	0.89	0.84	125
Crime	0.65	0.61	0.63	69
Economic	0.71	0.71	0.71	76
Failure	0.77	0.74	0.76	105
Terrorism	0.59	0.38	0.46	63
Social Aspects	0.70	0.79	0.74	160
Anxiety	0.71	0.77	0.75	150
Corruption	0.71	0.72	0.72	100
Poverty	0.77	0.72	0.71	80
Avg/Total	0.71	0.71	0.70	774

Table 1 is describing dataset description in which ten classes are described. Division of dataset where training and testing dataset is described. The second table is showing Logistic regression testing accuracy in which precision, recall, fl-score and support is described. The logistic regression test accuracy is calculated as 0.7144% which is almost 70 percent. The results clearly depicts that logistic regression has shown better performance while detecting the emotions.

Conclusion

These classes are extracted from these three tags, in this research we only focused on these eight classes according to their tags. These are eight different classes for Anger, Fear, Sadness, these are the main emotions from which you can define a person's behavior according to any situation, for example, if a person is angry he will be showing anger to that community which is the reason for his/her angry behavior.

Such as in social media person posts images of poor and needy people, definitely anyone who is commenting or tweeting on that image will be sad or angry according to this situation showing their emotions toward poor and needy peoples. The dataset used in this research is basically from areas in Pakistan to identify the reasons affecting the people or community of Pakistan, by applying different techniques and algorithms of NLP we designed models and applied those models on our dataset.

After analyzing we observed that there are eight factors that are affecting the people and community of Pakistan. These factors can be useful for reducing the affection of community and the main reason which is affecting especially the Pakistani community is Politics. In this research different approaches and techniques were carried out for analyzing the dataset after analyzing different models were adopted according to their accuracies and their results were compared. Furthermore, these techniques can be applied to identify the reasons of affecting the community in the overall globe. Identifying the reasons affecting the community can be used to reduce the tension, anxiety, fear, anger, and also some more known disturbing human health. We can also use these reasons to induce the amount of happiness, satisfaction, and pleasure because from a happy mind we can attain better health.

The dataset used in this paper is limited, twitter Pakistan dataset is limited, moreover the model can further be combined with other NLP algorithm to find better accuracy. Moreover, the analysis of the sentiments by using specific keyword is limited. For this we can use opinion lexicon method for improving the accuracy of the model. We can further focus on narrowing down the analysis to positive side of sentiments as well and use different domains and different languages.

Author's Contribution: B. J., Conceived the idea; C.B., & A.M., Designed the simulated work and K.A., & A.M., did the acquisition of data; C.B., & B.M., Executed simulated work, data analysis or analysis and interpretation of data and wrote the basic draft; C.B., & A.M., Did the language and grammatical edits or Critical revision.

Funding: The publication of this article was funded by no one.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgement: The authors would like to thank the venerated professors of Computer Science and IT department University of Balochistan, Quetta for assistance with the collection and analysis of data.

REFERENCES

- [1] K. Sailunaz, & Alhaji, R. , "Emotion and sentiment analysis from Twitter text," *Journal of Computational Science*, vol. 36, 2019.
- [2] M. Koper, Kim, E., Klinger, R., "Ims at emoint-2017: Emotion intensity prediction with affective norms, automatically extended resources and deep learning," in *In: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2017*, pp. 50-57.
- [3] S. Kiritchenko, Mohammad, S., Salameh, "M.: Semeval-2016 task 7: Determining sentiment intensity of english and arabic phrases," in *In Proceedings of the 10th international workshop on semantic evaluation (SEMEVAL-2016), 2016*, pp. 42-51.
- [4] Y. He, Yu, L.C., Lai, K.R., Liu, "W.: Yzu-nlp at emoint-2017: Determining emotion intensity using a bi-directional lstm-cnn model," in *In Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2017*, pp. 238-242.
- [5] M. Biswas, "SENTRAC: A Novel Real Time Sentiment Analysis Approach Through Twitter Cloud Environment," in *First International Conference on Advances in Electrical and Computer Technologies 2019 (ICAECT 2019), At: Hotel Aloft, Coimbatore, Tamil Nadu, 2020*.
- [6] A. Sharma, & Ghose, U. , "Sentimental Analysis of Twitter Data with respect to General Elections in India," in *Procedia Computer Science 2020*, pp. 325–334.
- [7] U. Sapkota, Steven Bethard, Manuel Montes, and Thamar Solorio. , "Not all character n-grams are created equal: A study in authorship attribution," in *In Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies, 2015*, pp. 93-102.
- [8] K. Saglani, & Janwe, N. , "Machine learning based sentiment," *International Journal of Emerging Technologies and Innovative Reseach*, vol. 7, pp. 425–428, 2019.
- [9] S. Kulkarni, & Kedar, P. , "A survey on twitter sentiment analysis," (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, vol. 5, pp. 22-25, 2020.
- [10] A. Kanavos, Vonitsanos, G., Mohasseb, A., & Mylonas, P., "An Entropy-based Evaluation for Sentiment Analysis of Stock Market Prices using Twitter Data," in *15th International Workshop on Semantic and Social Media Adaptation & Personalization, 2020*.
- [11] Z. Dai, Chenyan Xiong, Jamie Callan, and Zhiyuan Liu. , "Convolutional neural networks for soft-matching n-grams in ad-hoc search," in *In Proceedings of the eleventh ACM international conference on web search and data mining, 2018*, pp. 126-134.
- [12] A. Kumar, Vikrant Dabas, and Parul Hooda, "Text classification algorithms for mining unstructured data: a SWOT analysis," *International Journal of Information Technology*, vol. 12, pp. 1159-1169, 2020.
- [13] J. Samuel, Myles, R. & Kashyap, R., "That Message Went Viral?! Exploratory Analytics and Sentiment Analysis into the Propagation of Tweets," in *In 2019 Annual Proceedings of Northeast Decision Sciences Institute (NEDSI) Conference, Philadelphia, USA, 2019*.
- [14] S. Singh, & Pareek, A. , "Improving Public Transport Services using Sentiment Analysis of Twitter data," *Journal of Information and Computational Science*, vol. 10, pp. 234–250.
- [15] S. Moriya, and Chihiro Shibata, "Transfer learning method for very deep CNN for text classification and methods for its evaluation," in *In 2018 IEEE 42nd annual computer software and applications conference (COMPSAC), 2018*, pp. 153-158.
- [16] B. A. V. K. N. Rajput, "Word Frequency and Sentiment Analysis of Twitter messages during Coronavirus Pandemic," vol. *CoRR*, Vol.abs/2004.03925, 2020.

- [17] A. Alazba, Alturayef, N., Alturaief, N., & Alhathloul, Z. , "Saudi stock market sentiment analysis using twitter data saudi," in International Conference on Knowledge Discovery and Information Retrieval – KDIR, 2020.
- [18] P. Fornacciari, Mordonini, M., & Tomaiuolo, M. , "Social Network and Sentiment Analysis on Twitter : Towards a Combined Approach," in 1st International Workshop on Knowledge Discovery on the WEB (KDWeb), 2015, pp. 1-2.
- [19] Rosenthal-etal-2017-Semeval.2017, "SemEval-2017 Task 4: Sentiment Analysis in Twitter " in Proceedings of the 11th International Workshop on Semantic Evaluation Association for Computational Linguistics, 2017.
- [20] G. A. Ruz, Henríquez, P.A., & Mascareño, A. , "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers," *Future Gener. Comput. Syst.*, vol. 106, pp. 92-104, 2020.
- [21] J. Samuel, Garvey, M., & Kashyap, R., "That message went viral?! exploratory analytics and sentiment analysis into the propagation of tweets," vol. arXiv preprint arXiv:2004.09718, 2020.
- [22] T. Pranckevičius, and Virginijus Marcinkevičius. , "Application of logistic regression with part-of-the-speech tagging for multi-class text classification," in In 2016 IEEE 4th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), 2016.
- [23] S. M. H. Dadgar, Mohammad Shirzad Araghi, and Morteza Mastery Farahani, "A novel text mining approach based on TF-IDF and Support Vector Machine for news classification," in In 2016 IEEE International Conference on Engineering and Technology (ICETECH), pp. 112-116.
- [24] K. Were, Dieu Tien Bui, Øystein B. Dick, and Bal Ram Singh, "A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape," *Ecological Indicators*, vol. 52, pp. 394-403, 2015.
- [25] I. A. Tamposis, Konstantinos D. Tsirigos, Margarita C. Theodoropoulou, Panagiota I. Kontou, and Pantelis G. Bagos. , "Semi-supervised learning of Hidden Markov Models for biological sequence analysis," *Bioinformatics*, vol. 35, pp. 2208-2215, 2019.
- [26] M. Hammad, and Haris Anwar, "Sentiment Analysis of Sindhi Tweets Dataset using Supervised Machine Learning Techniques," in In 2019 22nd International Multitopic Conference (INMIC) 2019.
- [27] V. A. Fitri, Rachmadita Andreswari, and Muhammad Azani Hasibuan, "Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm," *Procedia Computer Science*, vol. 161, pp. 765-772, 2019.
- [28] M. F. M. J. Shamrat, Sovon Chakraborty, M. M. Imran, Jannatun Naeem Muna, Md Masum Billah, Protiva Das, and O. M. Rahman, "Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm," *Indonesian Journal of Electrical Engineering and Computer Science* vol. 23, pp. 463-470, 2021.
- [29] Hammad M, Anwar H. "Sentiment Analysis of Sindhi Tweets Dataset using Supervised Machine Learning Techniques." In 2019 22nd International Multitopic Conference (INMIC), pp. 1-6. IEEE, 2019.
- [30] K. Chowdhary, "Natural language processing," In *Fundamentals of artificial intelligence*, pp. 603-649, 2020.