

A Review: Sentiment analysis methods and their use in social media platforms

TEJA VENKATA SAI OBULAPURAM
Department of computer science
and Engineering,
Koneru Lakshmaiah
Education Foundation,
Vaddeswaram, Andhra Pradesh,
India
tejaobulapuram2003@gmail.com

JASWANTH SAI REDDY YENDREDDY
Department of computer science
and Engineering,
Koneru Lakshmaiah
Education Foundation,
Vaddeswaram, Andhra Pradesh,
India
jsai2099@gmail.com

LAKSHMI SARVAGNA KOTARI
Department of computer science
and Engineering,
Koneru Lakshmaiah
Education Foundation,
Vaddeswaram, Andhra Pradesh,
India
lakshmisarvagnakotari@gmail.com

VARUN TEJA TIRUMALASETTY
Department of computer science
and Engineering,
Koneru Lakshmaiah
Education Foundation,
Vaddeswaram, Andhra Pradesh,
India
t.varunteja123@gmail.com

SANTHI SRI.T
Department of computer science
and Engineering,
Koneru Lakshmaiah
Education Foundation,
Vaddeswaram, Andhra Pradesh,
India
sri_santhi2003@yahoo.com

S.SAGAR IMAMBI
Department of computer science
and Engineering,
Koneru Lakshmaiah
Education Foundation,
Vaddeswaram, Andhra Pradesh,
India
simambi@gmail.com

Abstract— The method for recognizing positive, neutrality, or negativity sentiment in text is known as sentiment analysis. Sentimental analysis is mostly helpful for enhancing customer service, promoting goods and services, anticipating opinion polls, enhancing awareness of the signification of data security, and identifying how spectators feel about a certain sport. This paper examines the methodology employed in a review of sentimental analysis across several platforms, with a focus on social media, and discusses its applications. Social media networks have a lot of unprocessed data that users post as text and video. The utilization of sentimental analysis can transform the data into useful knowledge. The following reliable sources, including IEEE Xplore, Science Direct, and Springer, were utilized to carry out a systematic review of studies from 2016. 23 of the 50 articles that were originally and thoroughly reviewed have been chosen for the review process. The results demonstrate that most articles used lexicon- and machine-learning-based approaches to analyze text sentiment on social media platforms, with applications in industry, medical services, and safety.

Keywords: Machine Learning, social media, Sentimental Analysis

I. INTRODUCTION

For the time being social media is influencing the world. Social media is a computer-based technology that creates online communities and networks to facilitate the sharing of concepts, ideas, and knowledge. Businesses later adopted it as a method to interact with clients through a well-liked new channel of communication after initially using it to contact friends and family. More than 3.8 billion individuals worldwide use social media sites including Twitter [1,3,5,6], Facebook, Instagram, and a host of others. These locations undergo continuous modification. Globally, more than 3.8 billion individuals use social media. By 2023, forecasting tools predict that there will be 4 billion social media users worldwide. To build a following, a fan base, and interact with the culture surrounding their own brand, businesses are also using social media marketing to direct target their clients on smartphones and laptops. So, to promote their businesses they need to understand the user's viewpoint like what the

users are trying to present by their text, videos, photos etc. So, the information that is obtained by the businesses are in the raw form and does not process the data i.e. unprocessed data and data are to be processed and transform into useful data which will be more useful for the business organization and users.

This article is mainly concentrated on understanding the sentimental analysis which uses various kinds of techniques for the platform like Twitter [10,17,19,20] etc. For this we have considered various journals related from the year 2016 to current which uses various methodologies to look up the performance and accuracy of the model that uses various dataset that would be helpful for the business organizations and users. Sentimental analysis is a technique for text mining that finds and extracts data from sources, helping organizations to comprehend the public opinion of their brands, products, and services. It uses the Natural Language Processing (NLP) to extract and convert the raw text into the text that is useful. The sentimental analysis categorizes various textual sentiments as neutral, positive, or negative as shown in “Fig.1”. The fundamental objective of sentiment analysis is to comprehend other people's viewpoints on a text or contact communication that incorporates contextual polarity.

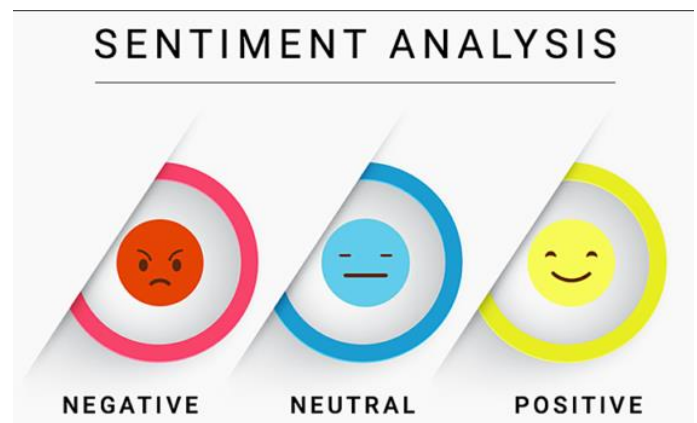


Fig.1. Example for neutral, negative, and positive sentiments

In general, the text expresses the facts and opinions, and the classification of the text can be done by using sentimental analysis. Facts represent objective expressions with regards to the evidence and the facts which are truth. Opinions represent subjective expressions with respect to the sentiments, emotions, feelings, and events. Sentiment analysis has recently employed significant ml and rule-based learning techniques. Research on text sentiment analysis, which is employed by data from the sentiment lexicon, has merged deep algorithm(dl) algorithms with machine learning(ml) algorithms. The sentiment lexicon approach yields the best outcomes. There are some main methods of sentimental analysis we can approach by using machine learning and lexicon-based [2,4,7,9] approaches (the number of articles that are used for this review based on the approaches are given in “Fig.2”). Lexicon-based approaches work by positive and negative count associated with the data, but machine learning approaches use various algorithms and the deep learning techniques to locate and collect the sentiment from the text.



Fig.2. Number of the articles and its approaches used for this review

However, in more recent times, sentiment lexicon models have also been able to reveal the hidden meaning of a sentence by using the parts of speech, and by doing so, we can enhance sentimental polarity and intensity values [2]. Creating a sentimental vocabulary that could distinguish between words that were neutral, positive, and negative was the main objective of sentiment lexicon-based models in the past. To determine the polarity of the sentiment, the terms in the input text are contrasted with those found in the sentiment lexicon [2,4]. The foundation of this analysis is the input text. This paper examined an overview of the methodology used to analyze each article's sentimental analysis method, the kind of platform utilized to collect information, and the application context used for the sentimental analysis.

II. LITERATURE WORK

In this the review process began with establishing the research objective, then choose the criteria that the research should have. Proceed by researching and choosing

appropriate materials from the pool of potentially relevant literature. One more procedure is to combine pertinent information from the research, and the last is to report the review's outcomes.

The following categories were examined to give an overview of the review. The method used to implement sentiment analysis on any platform, such as Twitter, the kind of platform used to examine each article's sentiment analysis technique, and how it may be applied to meet the needs of the public.

Many articles and research papers based on sentimental analysis from reputable sources, including IEEE, Science Direct, and Springer, were studied and evaluated. After examining the abstracts of more than 50 papers, we chose 25. 23 papers were ultimately chosen for examination, and it's provided in Table 1.

The sentimental analysis techniques are listed in Table 1 and are divided into the categories: based on lexical approaches, machine learning techniques or both. The result is then presented in Table 1 of each article based on the methodology employed. Table 1 additionally includes a column that includes the information used to perform the entire study or context.

III. DISCUSSION

A. Sentimental Analysis methods review context

According to the papers we analyzed, every publication showed how to implement sentiment analysis using either a method that relies on a lexicon or on machine learning, or a blend of each. The results suggest that, to do sentiment analysis, 4 of the publications that were analyzed used the method relies on a lexicon, 18 articles on research used method relies on Machine Learning in addition to 1 study demonstrated use of those two methods together [15] as shown in “Fig 2”.

Lexicon method approach is based on the learning technique that is unsupervised. It doesn't need to have any train data and it's according to the dictionary words. The study adapted the sentiment lexicons which are based on the intensity scores, Parts of Speech [4] while performing sentimental analysis. The adoption of this technique of is dependent on the frequency of the words in the textual data, whether these are positive, negative, or neural, in addition to the intensity score of each term with the predetermined polarity of lexicons. When using the Parts-of-Speech (POS) methodology, the sentiment lexicons are organized in accordance with the parts of speech and frequency of each text.

Lexicon techniques reliable resources, and the quality of such resources is critical for the overall technique's performance. It is focused on the polarity of the text, which may be identified by the polarity of the words that comprise it.

TABLE I: Summary of literature reviewed

S. No	Tool/Method	Result	Context/Based On
1	Machine Learning (SentiDiff) [1]	Improvements to the PR-AUC result	Twitter
2	Lexicon-based (related to intensity scores) [2]	Enhance the existing techniques for sentimental categorization using sentimental and conventional embeddings.	SemEval, Stanford Sentiment Treebank datasets
3	Machine Learning (Related to Linear Discriminant Analysis) [3]	Framework accurately detects changes in sentiment levels using an improved sentiment variation measurement.	Twitter
4	Lexicon-based (Based on Parts-Of-Speech) [4]	The constructed FCP-Lex beats older sentiment models with a high accuracy (over 80%) in both negative and positive corpora, according to the findings of text sentiment classification tests.	Long Text Review Dataset (Movie)
5	Machine Learning (related to CNN) [5]	This method produces embeddings on tweets, which were used as feature vectors in the CNN model of the input data	Twitter
6	Machine Learning (Based on Linear Discriminant Analysis) [6]	According to the research, the public preferred the lockdown and stay at home tactic based on the LDA algorithm	Twitter (COVIDSENTI)
7	Lexicon-based [7]	This gives the RGWE method based on the sentiment idea. They determine the best word sentiment idea for every situation and offer more precise word semantics and sentiment representation.	SemEval, Stanford Sentiment Treebank datasets
8	Machine Learning [8]	The results determines that the high level of effectiveness that this hybrid sentiment analysis method is capable of.	A well-balanced corpus website
9	Lexicon-based (Long Short-Term Memory) [9]	Two distinct situations are given to illustrate a novel method for determining the attention vector in general sentiment analysis without a goal. However, the attention mechanism only performs better on extended sequences in these two situations.	IMDB, Yelp2013 NB4000
10	Machine Learning [10]	Compared to all other algorithms, Decision Trees produce the best outcomes.	Twitter
11	Machine Learning (Based on Genetic Algorithm) [11]	By approximately 4, 0.8%, and 2 respectively, the accuracy of the Genetic Algorithm-based feature reduction technique surpasses the accuracy of the non-Genetic Algorithm-based feature selection.	UCI ML repository
12	Machine Learning [12]	Improved text representation of the remarks is possible thanks to the BiLSTM model's complete accounting for the context information.	Ctrip (Hotel Comment Text)
13	Machine Learning (Based on MCNN-MA model) [13]	In comparison to previous baseline models, the MCNN-MA model shows a higher classification accuracy and a comparatively cheap training time cost.	Two Chinese datasets: Tan Songbo's, Taobao Chinese
14	Machine Learning (Based on hybrid multi objective feature selection algorithm) [14]	The suggested model shows promise in terms of accuracy for improving sentiment classification performance in datasets from various areas.	Feature selection model
15	Machine Learning & Lexicon-based (On recommender systems) [15]	The purpose of rating prediction is accomplished by merging item reputation similarity, interpersonal sentiment effect, user sentiment similarity into a single matrix	User review corpus

		factorization framework.	
16	Machine Learning (Based on distantly supervised lifelong learning framework) [16]	It provides a comprehensive architecture that makes it possible to modify any mono task sentiment algorithm for the benefit of ongoing sentiment learning.	Chinese Weibo dataset & English Twitter dataset
17	Deep learning (LSTM, BiLSTM, and GRUs) [17]	BiLSTM offers the most accurate, precise, recallable, and experimental outcomes when compared to F1 LSTM and GRU models (F1).	Twitter
18	Machine Learning (Based on Convolution) [18]	When analyzing sentiment on Twitter, the deep convolutional neural network that uses word vectors that have already undergone training performs well.	The deep convolutional neural network model produced superior accuracy and F1-measure scores for classifying tweet sentiment.
19	Machine Learning (LR, NB, SVM, RF) [19]	According to experimental findings, URLs, stop words, and digits have little bearing on classifier performance, however that negation replacement and acronym lengthening can increase classification accuracy.	Twitter
20	Machine Learning (Based on multi-class classification) [20]	In contrast to the 81.3% accuracy of the binary classification, this method achieves an accuracy of 60.2% for each of the 7 separate sentiment classifications.	Twitter
21	Machine Learning [21]	It has three weighting algorithms and four distinct n-gram feature sets. In comparison to other feature sets, the unigram features are preferred due to their precision.	SemEval 2016 dataset
22	Machine Learning (Based on DL models like HAN, Convolutional Neural Network, Recurrent Neural Network) [22]	The DL algorithms outperformed the TML methods on the data that had been stemmed.	Traditional machine learning model
23	Machine Learning (Based on Naive Bayes) [23]	A classifier that uses rules that combines emoticons and words that reflect emotion with the unsupervised Naive-Bayes classifier is used to categorize the sentiment of tweets.	Stanford Twitter Sentiment140 dataset

This methodology is not made to address every aspect of language because natural languages are so complex, when it pertains to colloquialism, irony, and negation. Utilizing words that is sentiment is inadequate. There are several issues, which include that some words can mean different things depending on the context [4], that many words without sentimental connotations might equally represent a viewpoint, and that many words with emotive meanings may not truly be communicating anything. However, the lexicon-based approach does have some strengths of its own, notably simple counting of both positive and negative terms, versatility to other languages, and speed of analysis.

The machine learning techniques adopted for the study are under the categories of deep learning and supervised learning, and both require training data to be processed. Techniques for machine learning have included the Support Vector Machine [8,19] (SVM), the Naïve (Independent) Bayes model [8,19,23], and Genetic Algorithm [11]. Numerous machine learning models exist; however, these are the most widely used ones. When applied to well-formed text corpora, naive Bayes is beneficial, while support vector machines function well on low-shape corpora. Machine learning methods, however, struggle with data sets other than social media, which frequently contain a huge proportion of spelling errors and arbitrary length corpora. To enhance these methods, a substantial amount of training data must be accumulated. Moreover, using machine learning for analysis requires a lot of time, sometimes hours in sophisticated models, especially if training is essential. A smaller training dataset speeds up the procedure, but the classification accuracy suffers.

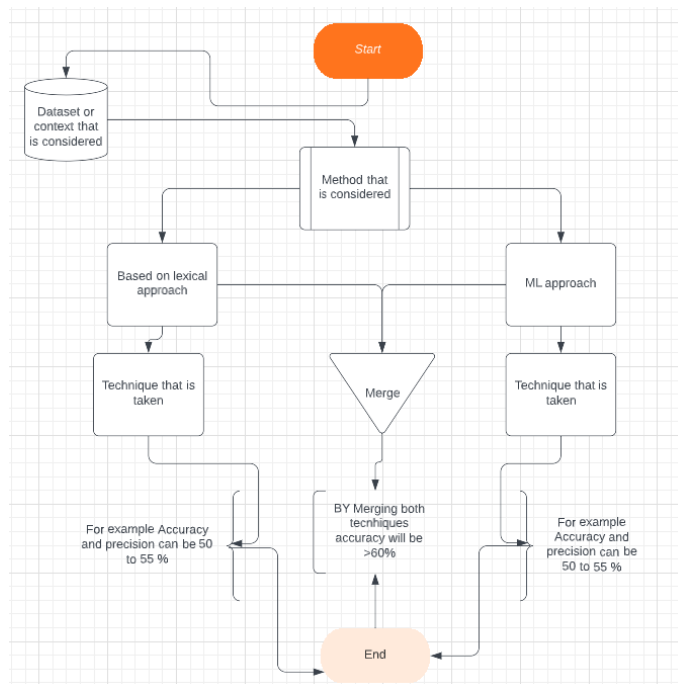


Fig .3. Flowchart for the process of each method that is considered

It's interesting to note that the accuracy performance of the Lexicon-based technique and the Machine Learning approach is remarkably similar. Combining two methods [15]—based on lexicon sentiment classification using scoring of sentiment functions and from a machine-learning perspective, Naive Bayes event models allow one to estimate the sentiment's direction. Studies have shown that using both

strategies together rather than just one improves performance (for example like shown in “Fig.3”). Therefore, combining the two ways is suggested to improve the results because they will complement one another and produce a better outcome than using one approach exclusively. When identifying a phenomenon, it can be helpful to combine two methodologies. It also makes it easier to process unstructured data.

B. Sentiment analysis using social media data collections context

Four types of social media exist: blogs - Reddit, content communities - Instagram, YouTube, microblogging - Twitter [24] and social networking sites - Facebook, LinkedIn.

Twitter is most popular social media network for obtaining information on user opinion, according to the study under consideration, among the four categories of social media services. In 85% of the studies, we looked at, Twitter was used to collect data for sentiment analysis. Users can post and exchange brief messages on Twitter, one of the ten most popular websites. Twitter is used by academics, businesses, and even the government to express their opinions and share important information.

Twitter is a well-known social media platform for microblogging where the users can share their thoughts about a particular person, situation, or item. With the use of APIs, users can locate and copy data on a certain subject based on keywords. Twitter has 500 tweets each day and makes its data available to the public via an API [25], allowing for in-the-moment analysis. Additionally, the official Twitter account of London Heathrow Airport [26] was utilized to collect tweets from users and analyze them for the UK energy firms [27].

Using SemEval with Stanford Sentiment Treebank datasets [2,7,21,23] are also very well-liked for sentiment analysis due to disorganized form of data. It is badly organized and filled with various spelling errors and short forms. Data analysis may become more challenging as a result.

C. Sentiment analysis application context

Sentiment analysis has several uses that can aid in decision-making in a variety of different fields. Sentiment analysis also has the benefit of increasing public knowledge of data security issues and security breach risks. It serves as an example of how companies should respond to security breaches to influence the public [25]. Sentiment analysis was also applied to the jobless rate and job growth sentiment score on social media. [28]. Sentiment analysis can aid when a disaster strikes by figuring out what people's emotional needs are so that a successful rescue effort can be planned [29]. To evaluate a person's level of melancholy, sentiment analysis also uses text to extract, overserve, and analyze emotions.

Sentiment analysis benefits business owners by enabling them to assess client satisfaction with their goods or services[30], as well as how effectively they interact with them on social media and how well their brand is doing in general[31].When sentiment analysis is used, customer

feedback is of vital importance because it may aid businesses and organizations in taking the necessary actions to enhance their products or services and business strategy effectively locate and react to emergencies. This demonstrates the crucial value of sentiment analysis for comprehending people's perceptions and nurturing in making decisions. Additional research is essential to devise a model of sentiment analysis that is usable to analyze the data categories besides English language, expand the contexts in which sentiment analysis is used, and provide recommendations for future.

IV. CONCLUSION

This paper summarizes the findings of studies on sentiment analysis that were undertaken in a systematic manner. The following are the contributions that the paper makes. We first demonstrate the technique for sentiment analysis. The most broadly applied methods of the Lexicon-based Method are Intensity Scores and Parts of Speech, whereas for Machine Learning, Support Vector Machine, Naive Bayes, and Genetic Algorithm are being used. The data alone determines the most appropriate sentiment analysis approach to implement. Both approaches exhibited similar accuracy. If the data structure is disorganized or there is little data, it is advisable to opt the lexicon-based approach, and the time you are given to perform analysis is tight. Techniques that rely on machine learning benefit from relatively large data because they require more training sets and training time. Integrating the lexicon-based method and the machine learning(ml)based strategy is suggested to enhance precision and the efficiency of the outcomes.

Next, we determine that Twitter is the best source for information retrieval. Most of the research papers that were analyzed use mainly Twitter as their context from the social networking media. Because Twitter content is widely available, easily obtainable, and content rich. Finally, we present a sentiment analysis application. Sentiment analysis has the plethora of usage and can be used to deliver enhanced quality and strategy in business related purposes, surmise the outcome of polls, detect disease outbreaks, create more awareness of the value of security and privacy, comprehend how individuals are feeling about a sport that is, and effectively identify and react to the emergencies. This exemplifies how sentiment analysis is vitally crucial to comprehending public perspective and for trying to foster in making decisions. Additional research to create a usable sentiment analysis model to analyze data categories other than English language data is required., expand the contexts in which sentiment analysis is used, and provide recommendations for future.

REFERENCES

- [1] Lei Wang, Shui Yu, Jianwei Niu., “SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis.” vol. 32, pp.2026 - 2039, April (2019)
- [2] K.Robert Lai, Xuejie Zhang, Liang-Chih Yu,Jin Wang., “Refining Word Embeddings Using Intensity Scores for Sentiment Analysis.” vol. 26, pp.671 - 681, December (2017)

- [3] Khaleed Shaalan, Fuad Alattar., "Using Artificial Intelligence to Understand What Causes Sentiment Changes on Social Media." vol. 9, pp. 61756 - 61767, April (2021)
- [4] LISHA LIN, FULLIAN YIN, JIANBO LIU, YANYAN WANG., "The Construction of Sentiment Lexicon Based on Context-Dependent Part-of-Speech Chunks for Semantic Disambiguation." vol. 9, pp. 63359 - 63367, March (2020)
- [5] Huyen Trang Phan, Do Sam Hwang, Van Cuong Tran, Ngoc Thanh Nguyen., "Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model." vol. 8, pp. 14630 - 14641, January (2020)
- [6] Matloob Khushi, Imran Razzak, Jinman Kim, Usman Naseem., "COVIDSenti:A large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis." vol. 8, pp. 1003 - 1015, January (2021)
- [7] YA ZHOU, GUIMIN HUANG, YABING WANG, JUN LI, HUI LI, HUA JIANG., "Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis." vol. 9, pp. 37075 - 37085, March (2021)
- [8] Ying Fang, Jun Zhang, Hai Tan., "Multi-Strategy Sentiment Analysis of Consumer Reviews Based on Semantic Fuzziness." vol. 6, pp. 20625 - 20631, April (2018)
- [9] JIANQIANG LI, XIANGHUA FU, MIN FANG, JINGYING YANG, HUIHUI WANG., "Lexicon-Enhanced LSTM With Attention for General Sentiment Analysis." vol. 6, pp. 71884 - 71891, October (2018)
- [10] Saad, & Yang, J., S. E., "Twitter Sentiment Analysis Based on Ordinal Regression." vol. 7, pp. 163677 - 163685, November (2019).
- [11] Iqbal, F., J., Fung, Maqbool, R., Batool, P.C. K., Khattak, Aleem, B. C. M., S., & Hung, A. M., "A Hybrid Framework for Sentiment Analysis using Genetic Algorithm based Feature Reduction." vol. 7, pp. 14637 - 14652, January (2019).
- [12] Xu, Y., Meng, X., Yu, G., Z., Wu, X., Qiu., "Sentiment Analysis of Comment Texts Based on BiLSTM." vol. 7, pp. 51522 - 51532, April (2019).
- [13] YAN CHENG, YUE FENG., "Short Text Sentiment Analysis Based on Multi-Channel CNN With Multi-Head Attention Mechanism." vol. 9, pp. 19854 - 19863, January (2021)
- [14] Merih Angin, Ayça Deniz, Pelin Angin., "Evolutionary Multiobjective Feature Selection for Sentiment Analysis." vol. 9, pp. 142982 - 142996, October (2021)
- [15] Xiaojiang Lei, Guoshuai Zhao, Xueming Qian., "Rating Prediction Based on Social Sentiment From Textual Reviews." vol. 18, pp. 1910 - 1921, June (2016).
- [16] Jie Jiang, Rui Xia, Huihui He., "Distantly Supervised Lifelong Learning for Large-Scale Social Media Sentiment Analysis." vol. 8, pp. 480 - 491, November (2017).
- [17] S.-L., Cheng, Tsai, L.-C., "Deep learning for automated sentiment analysis of social media." April (2020)
- [18] Gui Xiaolin, Zhao Jianqiang, Zhang Xuejun., "Deep Convolution Neural Networks for Twitter Sentiment Analysis." vol. 6, pp. 23253 - 23260, January (2018).
- [19] Z., Jianqiang, G., Xiaolin., "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis." vol. 5, pp. 2870 - 2879, February (2017).
- [20] Tomoaki Ohtsuki, Mondher Bouazizi., "Multi-Class Sentiment Analysis on Twitter: Classification Performance and Challenges." vol. 2, pp. 181 - 194, April (2019).
- [21] A., Sharan,Naz, Malik, N.,S., "Sentiment Classification on Twitter Data Using Support Vector Machine." January (2018).
- [22] Md. Haris Uddin Sharif, Sezai Tokat, Harisu Abdullahi Shehu, Rabie A. Ramadan, Md. Haidar Sharif, Sahin Uyaver, Ripon Datta, Huseyin Kusetoğullari., "Deep Sentiment Analysis: A Case Study on Stemmed Turkish Twitter Data." vol. 9, pp. 56836 - 56854, April (2021).
- [23] Chy, Siddiqua, Ahsan, A. N., U. A., T., "Combining a rule-based classifier with weakly supervised learning for twitter sentiment analysis." February (2016).
- [24] Ali, Kashif, Athman Bouguettaya, Hai Dong, Abdelkarim Erradi, and Rachid Hadjidj., "Sentiment Analysis as a Service: A Social Media Based Sentiment Analysis Framework", (2017)
- [25] Hao,Hongying Dai and Jianqiang., "Social Media Content and Sentiment Analysis on Consumer Security Breaches." Journal of Financial Crime 23 (2016)
- [26] Martin-Domingo, Juan Carlos Martin, Luis, and Glen Mandsberg., "Social Media as a Resource for Sentiment Analysis of Airport Service Quality (ASQ)." (2019)
- [27] Ikoro, Maria Sharmina, Victoria,Riza Batista-Navarro and Khaleel Malik., "Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers"(2018)
- [28] Phoong Seuk Wai,Shahid, Ainin Sulaiman,Yeong Wai Chung, Noor Ismawati Jaafar, Shayaa and Shamshul Bahri Zakaria., "Social Media Sentiment Analysis on Employment in Malaysia", (2017)
- [29] Ragini, P. M. Rubesh Anand, J. Rexiline and Vidhyacharan Bhaskar., "Big Data Analytics for Disaster Response and Recovery Through Sentiment Analysis." International Journal of Information Management (2018)
- [30] Rahman,F. A. AlOtaibi, S. A. El, and W. A. AlShehri., "Sentiment Analysis of Twitter Data", (2019)
- [31] Poecze,Claus Ebster, Flora, and Christine Strauss., "Social Media Metrics and Sentiment Analysis to Evaluate the Effectiveness of Social Media Posts." (2018)