



# INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

## A RESEARCH STUDY ON DIFFERENT MACHINE LEARNING TECHNIQUES FOR SENTIMENTAL ANALYSIS

*Baby Sudha. S,  
Research Scholar,  
Sri Krishna Arts and Science College,  
Coimbatore.*

*Dr. S. Dhanalakshmi,  
Assistant Professor,  
Sri Krishna Arts and Science College,  
Coimbatore.*

### ABSTRACT

Many organizations that directly reach the public using social media networks to analyse their brand value as well as the quality of services provided by them. It connects with customers, and collects information about their thoughts and views. Sentiment analysis is a machine learning technique that detects polarities such as positive or negative thoughts within a text, complete document, paragraph, line, or subsection. Machine learning (ML) is an interdisciplinary field, a hybrid of statistical and computer science algorithms commonly used in predictive analytics and classification analysis. This paper presents common sentiment analysis approaches and sentiment analysis techniques from a machine learning perspective. Some of the widely used prominent machine learning algorithms for opinion mining / Sentiment analysis like Naïve Bayes, Support Vector machine, Logistic regression, linear regression, Decision Tree, etc. Sentiment Analysis is a booming trend which is used in variety of applications such as social media monitoring, brand and reputation management, product analysis, market research, stock market analysis, human computer interaction, etc.

**Keywords:** Sentiment classification, Sentiment analysis, Machine Learning algorithms.

### I. INTRODUCTION:

Twitter, Facebook, Instagram and WhatsApp like social media files are great tool to get connected with large number of people at the same time virtually which is not possible physically. They open the door to share one's opinion, feelings about the products, services, movies, etc.(Yi & Liu, 2020)[1]. The data is a factor which drives economy. It is important for both the customers and business organisations. These days people buy the product through online only based on the review or feedback provided by the customers. E-commerce sites and all the online service providers encourage the trend of posting comments about their product and services. Online product sellers and service providers believe the comments will helps to improve their sales and they used it to modify their product and services to meet the customer needs. Sentiment Analysis or Opinion Mining is the research area which is essential in the modern world (Vohra & Teraiya, 2013)[2]. Sentiment analysis can be done with the help of machine learning. It explores the trends and patterns hidden in the data. Machine learning models can be used to analyse the large volume of data easily and accurately which is tedious process for the human being. It is also a big deal to use correct algorithm with required proper parameter values.(Machine Learning & its Applications Outsource to India, 2020)[3].

### APPROACHES FOR SENTIMENT ANALYSIS

#### Lexicon-based approach:

Assess the polarity of your opinion using a dictionary that contains both positive and negative terms used in lexicon. The number of optimistic and pessimistic words is explained in the text. If the sentence contains the non-negative word then positive score will be allocated. If the text contains a large number of negative or pessimistic words, the text will receive a negative score. While the text has equal number of positive and negative words, then it will be marked as (zero) neutral. A lexicon of opinion (positive and negative opinions) is developed to finalize the word is positive or negative. There are several ways to build and compile a dictionary (Medhat et al., 2014)[5].

**Dictionary based approach:**

A small number of words of opinion with established guidelines are gathered manually (Medhat et al., 2014)[5]. In WordNet, SQuAD like corpus has collection of words are then searched and appended. It reduces gradually until there are no new terms. This approach depends heavily on the dictionary size, and the strength of the sentiment classification. As the dictionary size increases, this approach is wrong (Jain & Dandannavar, 2016)[4].

**Corpus based approach:** It has enormous corporations for semantic and syntactic opinion patterns. The collection of words in the corpus are context oriented and it requires the huge amount of labelled data type (Jain & Dandannavar, 2016)[4].

**Machine learning based approach:**

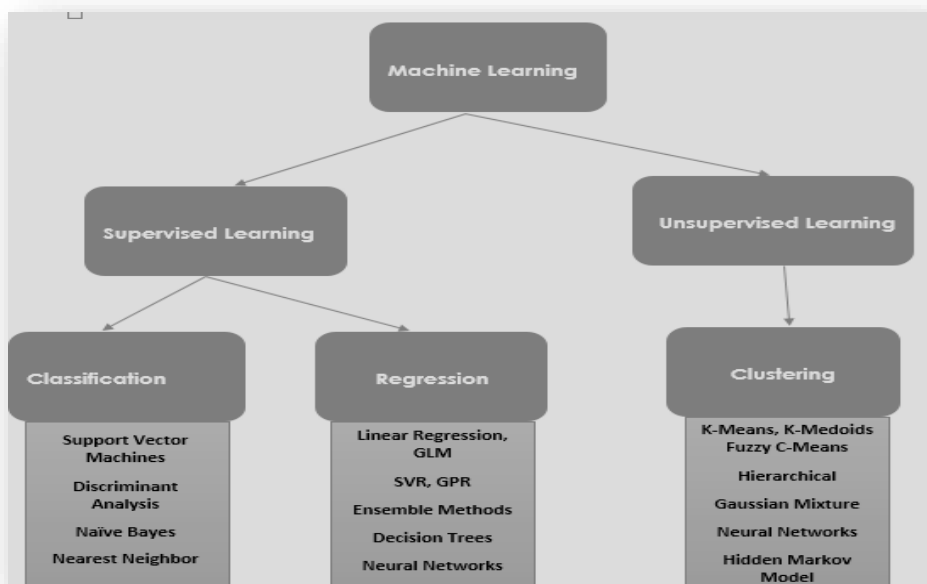
It applies machine learning algorithms for text classification. The classification of the sentiment based on machine learning can be categorized primarily into supervised and unsupervised methods of learning (Aydogan & Akcayol, 2016)[6].

**i)Supervised learning:**

A supervised learning technique depends on labelled data. This is a trusted and widely used method for sentiment classification. Frequently used methods are Naïve Bayes, Support Vector Machine, Artificial Neural network, Decision Tree, etc. Some less commonly used algorithms are Logistic Regression (LR), K-Nearest Neighbor (KNN), Random Forest (RF), and Bayesian Network (BN) (Aydogan & Akcayol, 2016)[6].

**ii)Unsupervised learning:**

This technique does not use prelisted data to train the classifier, unlike supervised learning. K-Means and Apriori algorithms are more common instance of the type of unsupervised machine learning algorithms. Unsupervised machine learning may also be divided into clusters and associations (Ahmad et al., 2017)[7].



**Fig 1: Approaches of Machine learning**

**Hybrid based approach:**

Hybrid based approach is the mixture of machine learning and lexicon based approach. Mostly Hybrid based approach is yields better research than other approaches. This hybrid approach is mainly beneficial as it can attain the top of both. The combination of Lexicon and Learning has demonstrated increased accuracy (Jain & Dandannavar, 2016)[4].

**II.BACKGROUND STUDY**

In general opinion surveys in the early 20th century, the science of sentiment analysis and opinion surveys has a strong foundation. Eventually, when online product reviews became necessary and accessible in the mid-2000s, they became a major research topic. In 2005, only 101 articles were published on this subject, but in 2015, about 5,699 articles were published. This means that sentiment analysis has grown almost 50 times in 10 years and has become one of the fastest growing areas of research in recent years (Mäntyläetal., 2018) [8]. In the early days of the Internet, people could get feedback from friends, neighbours and relatives before making a decision. Opinions sampling, research, and public opinion about the product or service were conducted by the organization. In recent year`s mobile devices, stock markets, and human emotions were other topics that have become popular (Mäntylä et al., 2018)[8]. Some of the challenges such as parallel computing of large amounts of data, irony, non-grammatical words are checking author segmentation, handling noise and dynamics. (Patil & Atique, 2015) [12].

Consumers can compare products based on people's reviews of those products. To make this more successful, they have developed supervised consumer assessment techniques. Describes two methods for classifying product characteristics according to consumer needs: the association rule method and the naive Bayes classifier. Not only is this analysis based on reviews, but personality is also important, and this sentiment analysis compares and identifies the preferred products that make consumers comfortable. The empirical assessment referred to two classifiers, the naive Bayes classifier and the class association rules (Yang

et al., 2010) [13]. Emotion reviews are accurately categorized by machine learning algorithms such as Bag of words, N-gram, Naive Bayes Classifier, and Natural Language Processing. Second, user feelings are categorized as positive, neutral, and negative, and the main features of the product make the customer interested in that particular product. This work also reviews products for opinions in multiple languages, copies the shortcomings when mapping slang, ridicules and copies the opinions, and then provides a comparative opinion between the two products best. One of the anaphoric resolutions (Gopu & Swarnatha, 2017) [14].

Sentiment analysis uses natural language processing to classify and derive emotions from text in a natural way. So, for example, in the consumer sector, it has a wide range of uses. Transfer education has also emerged as a new machine learning method that uses existing knowledge to solve problems and generate predictive results. It also includes sentiment analysis prospects such as applying aspects of cross-domain layer learning that have not yet been fully considered, and it is very difficult to use transfer learning to solve negative text data problems. It is difficult. And they conclude that in the future, aspect-level sentiment analysis of small texts is considered to be the most promising research method. (R. Liu et al., 2019) [15]. Machine learning algorithms such as the Naive Bayes, Support Vector Machine, and Maximum Entropy Classifier algorithms can be used for sentiment analysis of large amounts of data. These techniques allow you to use large amounts of data to make optimized strategic decisions. Sentiment analysis, also known as opinion mining, analyses the emotions that users show on social media such as Facebook, Yammer, Twitter, and microblogging, and responds to them with a human response.

### III. OUTLINE OF MACHINE LEARNING TECHNIQUES

Machine learning is a branch of Artificial Intelligence, which is used to identify hidden knowledge from the given input. The machine learning techniques are applied to wide variety of fields like Natural language processing, computer vision tasks, data mining, search engine, medical image processing, credit card fraud detection, market analysis, speech recognition, robotics, games etc. (Patel, 2018)[17].

The below mentioned are some fashionable machine learning algorithms.

#### 1. Linear Regression:

This is the tradeoff between dependent and independent variables. In this method the value of the dependent variable is calculated using the independent variable. Simply this is the mapping of dependent and independent variables on a line and that line is called as regression line and given by the formula  $y = a + bx$ . where  $y$  is the dependent variable,  $x$  is the independent variable and  $a$  is the intercept  $b$  is the slope of the line.

**2. Logistic Regression:** This technique is used to define the dependent variable from the group of separate variables. It is based on estimating probabilities using a logistic regression formula.

**3. Decision Tree:** It is useful for both classification and regression tasks by building a tree like structure. The best suitable attributes of a dataset is taken for building a tree for the classification or regression task. Then it splits the training data in to several subsets. The decision trees can be used to predict the class (Classification) or value (regression).

**4. Support vector machine:** SVM is a binary classifier. It is a linear model applicable to both classification and regression tasks. It can be useful in the process of solving linear and nonlinear problems. The data can be classified into their relative groups with the help of a line or a hyper plane.

**5. Naive-Bayes:** The technique is based on the Bayes theorem. It predicts the membership probability, which means the probability for a given data belongs to a particular class. The corresponding class with the highest probability will be considered as the class of the data.

**6. KNN:** This is useful for both classification and regression tasks. It is also called as instance based algorithm. It works based on the principle of selecting the  $k$  value. It makes the prediction directly from the dataset used for training purpose.

**7. K-means Clustering:** This is the type of unsupervised learning algorithm. Here  $k$  represents a number. Centroid represents the imaginary or real center location in a cluster. The  $k$  defines the number of centroids the user requires in the dataset Cluster means groups. This is the iterative algorithm which tries to divide the dataset into  $k$  predefined unique non overlapping clusters. Here the data strictly belongs to only one group.

**8. Random Forest:** This belongs to supervised machine learning algorithm. This is prominently used for both classification and regression problems. It is based on the principle of building decision trees on the different groups of data. Classification task using this technique requires the majority of the vote and the regression task requires the average of voting.

**9. Dimensionality Reduction Algorithms:** The term dimensionality reduction refers to the process of reducing the number of input features in the training data. This algorithm is useful in visualization. Independent Component Analysis (ICA) is a method based on the information theory is also a popular dimensionality reduction algorithm.

**10. Gradient boosting and Ada Boost Algorithms:** The gradient boosting algorithm is used for both classification and regression. Gradient boosting builds a prediction model which is a combination of several weak classifiers. Ada boost is easy to implement and it gradually minimizes the mistakes of the weak classifiers and it used to improves the accuracy by combining several weak classifiers.

#### IV. APPLICATIONS OF SENTIMENTAL ANALYSIS USING ML TECHNIQUES

- Sentiment Analysis using NLP techniques – Sentiment analysis requires large volume of data. Suryawanshi et al., (2020) [18]. It is needed one to understand peoples thought. NLP approaches are used to gather the data from twitter and process them into a meaningful format that machine models can understand. Here processing can includes hash tag removal, link removal, reputation removal, correcting spellings, deleting stop words, unnecessary spaces, punctuation marks, etc. It can be done by vader.
- Twitter sentiment analysis using Multinomial Naïve Bayes and Logistic Regression - Sentiment analysis involves several steps beginning from the data collection, preprocessing and model building. . The following author uses the test set results by the Logistic Regression with Count Vectorizer features (*Sentami Iselvan K*, 2020)[19].
- A Sentimental Analysis in Multilingual Web Texts can be done using machine learning approach. Boiy & Moens (2009)[24]. The classification of tweets can be done by machine learning algorithms such as SVM, Naïve Bayes, decision tree, etc.
- It is a newly introduced method, such as the speech pieces. This is also a mixture of adverbs, adjectives, and verbs that best in corporate speech parts. The Usage of Sentiment Analysis for Hearing the Voice of the Customer and Improving Businesses The analysis of sentiment studied to categorize the customers by blogs phrases and sentences.
- Social media monitoring: Social media posts are one-sided and often contain the most honest opinions about products, services, and businesses. Customer support management: It deals with variety of product and customer queries. Sentiment analysis helps in understanding the customer queries with the intended emotions, meaning and the purity.
- Brand and reputation monitoring management: Brand monitoring is one of the most common uses of sentiment analysis in business. Bad reviews can snowball online, and things get worse the longer you go. Sentiment analysis tools instantly notify you of negative brand mentions.
- Listen to employees By analyzing employee feedback sentiment, employer can learn how to increase employee engagement, reduce turnover, and increase productivity.

#### SUMMARY OF RELATED WORK:

*Table1: Review of findings presented by different authors between 2012 and 2020.*

S.No.	Author(s)	Year	Techniques and Conclusion derived
1	Kumar,A.,& Sebastian,T.M. [9]	2012	Fundamental terms, tasks and granularities, functional and future applications and Challenges of emotions are examined.
2	Duric,A.,&Song, F.[26]	2012	Learn the features of feedback automatically and identify them into positive, negative and neutral comments.
3	Vohra,S.,&Teraiya[2]	2013	Exploring and explaining its policies in this field by contrasting the concept of emotional research in natural Language processing.
4	Medhat, W., Hassan, A., & Korashy,H.[5]	2014	Gives the overview and review of the latest SA algorithms and software updates.
5	Cao,J.,Ke,Z.,Wang,H., Cheng,J., Qiao,F., Wen, D., & Gao,Y.[25]	2014	Overview of applications of TSA, creation of similar bases for the TSA program, contrasting both rules-based and learning based approaches' advantages and disadvantages based on web data characteristics
6	Cruz Diaz, N., Taboada,M.,&Mitkov,R.[21]	2015	A lexical knowledge and syntactic functions is adequate to identify the reach of a keyword immediately recognizes the clues.
7	Ravi,K.,&Ravi,V.[23]	2015	Explores the significance and vagueness of the model using fluid logic of Neural Network, Support Vector Machine, and lexicon based approaches.
8	Aydogan, E., &Akçayol,M.A. [6]	2016	Machine-based sentiment analysis is analyzed, identified by their knowledge extraction tasks and the possible problems are addressed.
9	Jain, A. P., &Dandannavar,P.[4]	2016	In the text analysis system for twitter data using Apache sparks, Naïve and Decision Trees Master Learning Algorithms are utilized for sentimental

			analysis.
10	Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. [7]	2017	Machine Learning methods and methods for emotion analysis and classification are used for many experiments and studies.
11	Swathi, R., & Seshadri, R. [10]	2017	Understand recent work on Big Data Machine Learning, devices and systems used and proposed methods and spaces.
12	Mäntylä, M. V., Graziotin, D., & Kuutila, M. [8]	2018	The automatic text clustering with manual qualitative analysis is possible to conduct a computerized literature review. For 6,996 papers, a bibliometric study of sentiment can be performed.
13	Mahendran, N., & Mekala, T. [12]	2018	A review of the available research methods, the issue and the application in the field of sentiment analysis
14	Shirsat, V., Jagdale, R., Shende, K., Deshmukh, S. N., & Kawale, S. [16]	2019	Sentence-level study of news stories and blogs to measure the polarity of news and forum posts by Naïve Bayes, the vector supporters and the random forest classification.
15	Bansal, A., Gupta, C. L., & Muralidhar, A. [22]	2019	Analyze YouTube feedback using various classification tools, including the decision tree, K-Nearest Neighbors and vector supports for strong predictions
16	Yi, S., & Liu, X. [1]	2020	Apply Machine Learning algorithms for learning, analysis and classification on the basis of customer experience of product data and shop knowledge.
17	Yogi, T. N., & Paudel, N. [11]	2020	Comparative tests of classification algorithms (SVM, Logistic Regression) based on machine learning are tested for sentiment analysis using various measurement parameters such as range, range retrieval and F calculation in the three different datasets of different size

## V. CONCLUSION & FUTURE WORK:

This paper briefly discussed the various machine learning based techniques for sentiment analysis from 2012 to 2020 also gives a brief introduction about the application of sentiment analysis in various fields. The analysis and extensive study of the literature gives an insight in the field of sentiment analysis. In future this work can be extended with citing all the works related to the sentiment analysis in any of the particular application area like product review from amazon / yelp or any other E-Commerce sites, or feedback from the patients of the hospital to get the opinion about the hospital among the patients. In future domain specific data can be analyzed extensively to obtain the good knowledge. This work can be extended to comparing the various machine learning algorithms by applying them on a same dataset and using various parameters accuracy can be compared.

## REFERENCES:

- [1] Jain, A. P., & Dandannavar, P. (2016). Application of machine learning techniques to sentiment analysis. *Second International Conference on Applied and Theoretical Computing and Communication Technology (ICATccT)*, 1(1), 628–632. DOI: <https://doi.org/10.1109/ICATccT.2016.7912076>
- [2] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. DOI: <https://doi.org/10.1016/j.asej.2014.04.011>
- [3] Aydogan, E., & Akcayol, M. A. (2016). A comprehensive survey for sentiment analysis tasks using machine learning techniques. *2016 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, 1(1), 1–7. DOI: <https://doi.org/10.1109/INISTA.2016.7571856>
- [4] Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine learning techniques for sentiment analysis: A review. *International Journal of Multi-disciplinary Science and Engineering*, 8(3), 27–35.
- [5] Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27(1), 16–32. DOI: <https://doi.org/10.1016/j.cosrev.2017.10.002>
- [6] Kumar, A., & Sebastian, T. M. (2012). Sentiment Analysis: A Perspective on its Past, Present and Future. *International Journal of Intelligent Systems and Applications*, 4(10), 1–14. DOI: <https://doi.org/10.5815/ijisa.2012.10.01>
- [7] Swathi, R., & Seshadri, R. (2017). Systematic survey on evolution of machine learning for big data. *International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1(1), 204–209. DOI: <https://doi.org/10.1109/ICCONS.2017.8250711>
- [8] Kawade, D., & Oza, K. (2017). Sentiment Analysis: Machine Learning Approach. *International Journal of Engineering and Technology*, 9(1), 2183–2186. DOI: <https://doi.org/10.21817/ijet/2017/v9i3/1709030151>
- [9] Patil, H., & Atique, M. (2015). Sentiment Analysis for Social Media: A Survey. *International Conference on Information Science and Security (ICISS)*, 1(1), 14. DOI: <https://doi.org/10.1109/ICISSEC.2015.7371033>
- [10] Yang, C., Tang, X., Wong, Y. C., & Wei, C.-P. (2010). Understanding Online Consumer Review Opinions with Sentiment Analysis using Machine Learning. *Pacific Asia Journal of the Association for Information Systems*, 2(1), 73–89. DOI: <https://doi.org/10.17705/1pais.02305>
- [11] Gopu, M., & Swarnalatha, P. (2017). Analyzing customer sentiments using machine learning techniques. *International Journal of Civil Engineering and Technology (IJCIET)*, 8, 1829–1842.
- [12] Liu, R., Shi, Y., Ji, C., & Jia, M. (2019). A Survey of Sentiment Analysis Based on Transfer Learning. *IEEE Access*, 7(1), 85401–85412. DOI: <https://doi.org/10.1109/ACCESS.2019.2925059>
- [13] Naiknaware, B., Kushwaha, B., & Kawathekar, S. (2017). Social Media Sentiment Analysis using Machine Learning Classifiers. *International*

*Journal of Computer Science and Mobile Computing*,6(6),465-472.

- [14] Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machinelearningclassifiers.*Human-CentricComputingandInformationSciences*,7(32),1-7.DOI:<https://doi.org/10.1186/s13673-017-0116-3>
- [15] Yogi,T.N.,&Paudel,N.(2020).ComparativeAnalysisofMachineLearningBasedClassificationAlgorithmsforSentimentAnalysis.*InternationalJournalofInnovativeScience,Engineering&Technology*, 7(6),1-9.
- [16] Patel, A. (2018.). Machine Learning Algorithm Overview. Medium. Retrieved on May 18, 2020from<https://medium.com/ml-research-lab/machine-learning-algorithm-overview-5816a2e6303>
- [17] Godara,N.,&Kumar,S.(2020).TwitterSentimentClassificationusingMachineLearningTechniques.*Waffen-UndKostumkundeJournal*,11(8),10-20.DOI:010.11205.WJ.2020.V1117.05.100959
- [18] Shirsat, V., Jagdale, R., Shende, K., Deshmukh, S. N., &Kawale, S. (2019). Sentence LevelSentiment Analysis from News Articles and Blogs using Machine Learning Techniques. *InternationalJournalofComputerSciencesandEngineering*,1(1),12-17.DOI:<https://doi.org/10.26438/ijcse/v7i5.16>
- [19] Yadav, N., Arora, M. S., & Tech, M. (2020). The Performance of Various Supervised MachineLearning Classification Algorithms in Sentiment Analysis of Online Customer Feedback in RestaurantSector of Hospitality Industry. *International Journal for Technological Research in Engineering*, 7(11),1-12.
- [20] AroteRutuja S., Gaikwad Ruchika P., Late Samidha S., Prof. G. B. Gadekar(2020). OnlineShoppingwithSentimentalAnalysisforFurnitureShop,*InternationalResearchJournalofModernizationinEngineeringTechnologyandScience*,02(05),1-8.
- [21] Raju, S. M., & Tarif, A. M. (2020). Real-Time Prediction of BITCOIN Price using MachineLearningTechniquesandPublicSentimentAnalysis.RetrievedonMay2020fromDOI:<http://arxiv.org/abs/2006.14473>
- [22] Ardianto,R.,Rivanie,T.,Alkhalifi,Y.,Nugraha,F.S.,&Gata,W.(2020).SentimentAnalysison E-Sports for Education Curriculum Using Naive Bayes and Support Vector Machine. *Journal ofComputerSciencesandInformation*,13(2),109-122.DOI:<https://doi.org/10.21609/jiki.v13i2.885>
- [23] Imtiaz,M.N.,&Islam,M.K.B.(2020).IdentifyingSignificanceofProductFeaturesonCustomer Satisfaction Recognizing Public Sentiment Polarity: Analysis of Smart Phone Industry UsingMachine-Learning Approaches. *International Journal of Applied Artificial Intelligence*, 1(1), 1-17.DOI:<https://doi.org/10.1080/088839514.2020.1787676>
- [24] Duan, W., Yu, Y., Cao, Q., & Levy, S. (2015). Exploring the Impact of Social Media on HotelService Performance: A Sentimental Analysis Approach. *Cornell Hospitality Quarterly*, 57(3), 282-296DOI:<https://doi.org/10.1177/1938965515620483>
- [25] Buczak, A., &Güven, E. (2015). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. *IEEE Communications Surveys & Tutorials*, 18(2),1153-1176. DOI:<https://doi.org/10.1109/COMST.2015.2494502>
- [26] Das, S., Behera, R., Kumar, M., & Rath, S. (2018). Real-Time Sentiment Analysis of TwitterStreamingdataforStockPrediction.*ProcediaComputerScience*,132(1),956-964.DOI:<https://doi.org/10.1016/j.procs.2018.05.111>

