

# Identification And Classification Of Text Using Sentiment Analysis And Lstm Algorithm

**Monish Kumar K**

*Undergraduate Student  
Information Technology  
Meenakshi Sundararajan Engineering College  
Chennai, India*

**Dr. A. Kanimozhi**

*Head of Department  
Information Technology  
Meenakshi Sundararajan Engineering College  
Chennai, India*

**Aswin Krishna V**

*Undergraduate Student  
Information Technology  
Meenakshi Sundararajan Engineering College  
Chennai, India*

**Vignesh K**

*Undergraduate Student  
Information Technology  
Meenakshi Sundararajan Engineering College  
Chennai, India*

**ABSTRACT:** Social events worldwide often provoke varied commentary. Sentiment analysis, a form of natural language processing (NLP), assesses text sentiment as positive, negative, or neutral. This method leverages linguistic data to help businesses monitor brand sentiment in consumer feedback and grasp client expectations. By extracting insights from unstructured online content like emails, blog posts, and social media comments, businesses can glean valuable information using rule-based, automated, or hybrid sentiment analysis algorithms. Automated tools, using machine learning, learn from data, while lexicon-based systems apply predefined criteria. Hybrid approaches, employing LSTM and BERT techniques, enhance sentiment classification. Twitter, a major data source, facilitates sentiment analysis of tweets, and Amazon stands out as a top shopping site based on daily data and reviews. Techniques such as CNNs are used for sentiment analysis; LSTM has excelled in multidimensional text analysis, achieving 97% accuracy. Tokenization with BERT (Bidirectional Encoder Representations from Transformers) has been effective in this domain. Experiments often leverage datasets like those from Kaggle—specifically, Twitter and Amazon data. Future research could integrate REST APIs and web crawlers for real-time tweet analytics, enhancing the sophistication and timeliness of sentiment analysis methods tailored to the dynamic nature of online discourse.

**KEY WORDS:** Long arrangement data, sentiment analysis, tweet analytics, deep learning, CNN, multidimensional approach, Twitter, Amazon, BERT.

Date of Submission: 29-04-2024

Date of acceptance: 08-05-2024

## I. INTRODUCTION

Sentiment analysis relies on textual mining to extract subjective information that measures brand social sentiment and improves business strategies. Consumer reviews are very important for tracking the reputation of products and brands. Sentiment analysis aids in the classification of emotions as negative, positive, or neutral by using the vast amounts of data generated on social media sites like Twitter. Because they can capture long-term dependencies, methods like Recurrent Neural Networks (RNNs), and especially Long Short-Term Memory networks (LSTMs), perform exceptionally well in this domain. The development of computational linguistics, in particular Natural Language Processing (NLP), has made it possible for computers to comprehend and process human languages with greater efficiency. Professionals can improve the quality of their writing by using tools like Grammarly, which offer correction suggestions and enhance articulation. By selecting suitable vocabulary and intonation, machine learning algorithms help with sentiment analysis and improve content comprehension. By following syntactic conventions and deriving contextual meanings, NLP techniques—such as syntactic and disambiguation procedures—validate the meaning of text. Sentiment analysis of social media data, particularly on Twitter, offers insightful information about public opinion that is applicable to a wide range of industries. Support Vector Machines (SVM) and ensemble classifiers are two machine learning classification algorithms that work well for sentiment analysis tasks. These algorithms let businesses make decisions based on customer feedback. The COVID-19 pandemic has brought to light the value of sentiment analysis in gauging public opinion on a range of topics, such as social distancing and vaccination. Accurately extracting sentiments from textual data requires the use of deep learning techniques like recurrent neural networks (RNNs) and convolutional neural networks (CNNs). These models can achieve high accuracy in sentiment classification tasks by utilising publicly available datasets, which can help decision-makers during emergencies such as the COVID-19 pandemic. All things considered, sentiment analysis—which is fueled by NLP and machine learning techniques—offers insightful information about public opinion, helping businesses and decision-makers make wise choices and effectively address new trends and crises.

## II. METHODS

One metric to gauge polarity is sentiment score. The term "polarity" refers to the ability of a word, phrase, or paragraph to convey sentiment. Thus, the system can conclude categorically that sentiment analysis plays a crucial role in polarity classification. This score is regarded as a numerical evaluation for either the entire text or specific phrases. Sentiment analysis with a finer point: To obtain a more specialised polarity rate, smaller groups ranging from a high percentage of positivity to very few negativities are separated. A 5-star rating system is contrasted with an analysis derived from user feedback.

**Concept-based analysis:** It helps to explain specific textual characteristics and attributes. The process of identifying sentiment and its features and characteristics based on a topic is known as ABSA. Features and themes are interchangeable in terms of theme.

**Emotion Detection:** In addition to positive and negative emotions, specific emotions such as joy, surprise, rage, frustration, and grief can be identified.

**Distance-based analysis:** the data's opinions and facts are distinguished. Negative feedback is generated by replacing a battery on the internet. For instance, customer support might get in touch with the system to address issues; this should be encouraged. Sentiment analysis is one of the emerging technologies that can be clearly predicted in advance and will take the same place as consumer portals and software as solutions models. The system monitored workflow and provided an overview of corporate applications that are more widely used and recognisable. Problems and restrictions with the analysis are also looked into in the current system.

Figure 1 illustrates how sentiment analysis can be used to analyse diverse emotions in expressive texts. Product reviews, customer feedback, and survey responses are some of the terms that are most frequently used in a variety of contexts, such as social media enhancement, business service management, and customer feedback analysis. Examining the comments on the product's features and pricing, for instance, is one way to see what the outcomes of product reviews are.



**Figure 1 Types of Sentiment**

Public opinion is understood to be what the general public wants by concentrating on businesses and brand perception. Online reviews, which represent the viewpoint of 93% of consumers, are the deciding factor in purchasing decisions, according to the findings of a Podium survey. Users' willingness may decrease as a result of negative reviews. They neglect to verify whether the feedback is positive. Finally, in order to live with this restriction, they will select a different course. Companies that use reputation fidelity as a strategy for problem-solving and that pay for sophisticated feedback techniques can outline this. Sentiment analysis allows for the precise measurement and verification of people's willingness to support a company over an extended period of time.

The general perception of a brand and people's willingness to interact with it are greatly influenced by public opinion. A Podium survey indicates that 93% of consumers believe online reviews have an impact on their purchasing choices. If users read a few unfavourable reviews, they might be less inclined to give the system a chance. They are not going to look into the possibility that the remarks were accurate. They will go in a different direction. In this environment, companies that keep a close eye on their brand can respond quickly to issues and improve productivity through feedback. In the era of computers, this kind of research makes it possible to measure people's opinions about companies precisely.

One of the most amazing methods in linguistics (NLP) for assessing the energy, hostility, and neutrality of a given piece of material is content analysis, also known as evaluation mining. Realising customer demands and offering consumer assessment based on factors such as the organization's support in maintaining high-quality goods and services are the main objectives. This has primarily occurred with text-based data that comprises random and unstructured content. These superfluous data are gathered from a variety of social media platforms, including websites, weblog posts, emails, support tickets, web accesses, virtual entertainment channels, chats, and opinions. They are identified by this receptive language mechanism that helps businesses. Information processing by hand has its limitations. Consequently, it is replaced by computations that are rule-based. Furthermore, machine learning techniques are disclosed for determining the sentiment polarity in preprogrammed arrangements. Many natural sentiment analysis techniques exist, such as standpoint sentiment analysis, fine-grained sentiment analysis, sentiment identification, and anticipation analysis.

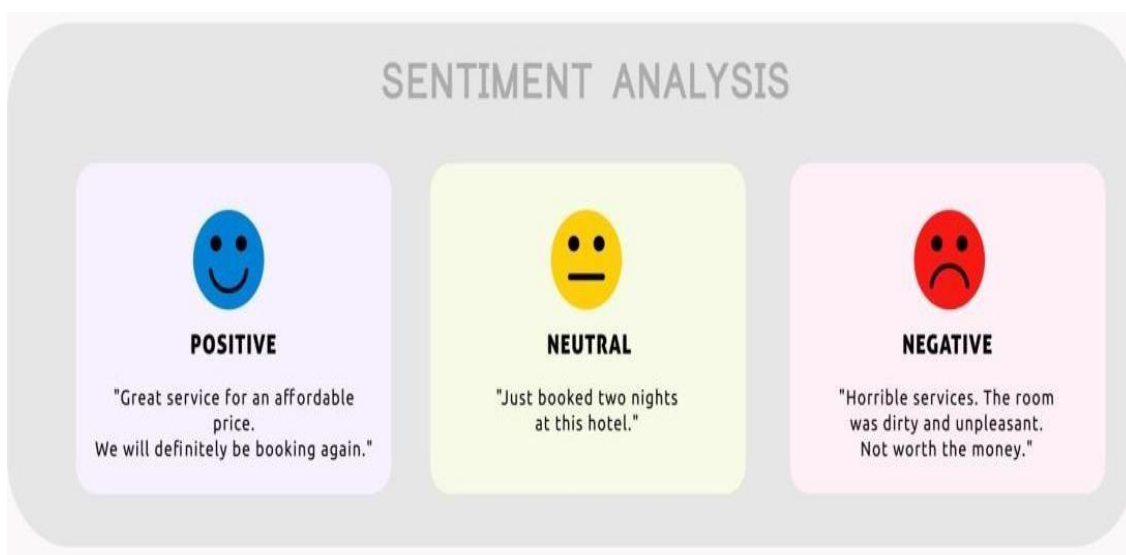
Polarity is measured by a parameter called sentiment score. The sentiment score is a numerical rating that can be applied to a paragraph's single phrase or the entire text. The term "polarity" refers to the ability of a word, phrase, or paragraph to convey sentiment. Thus, the system can state categorization of polarity plays a crucial role in Sentiment Analysis. Sentiment analysis needs can be matched with categories that the system can predict and adjust based on the explanation of customer reviews and research. Several popular techniques for sentiment analysis include:

**Sentiment analysis with fine detail:** Generally, the polarity shifts from a high percentage of positivity to a relatively small percentage of negativity. smaller group to obtain a more specialised rate of polarity. An analysis based on user opinions is compared with a 5-star rating system.

**Sentiment analysis based on aspects (ABSA):** This method is helpful in describing specific textual characteristics. ABSA stands for topic-based sentiment analysis and feature and characteristic finding. Themes and features are equivalent in terms of themes.

**Emotion Detection:** It is possible to identify specific emotions in place of positive and negative ones, such as joy, surprise, rage, frustration, and grief.

**Intent-based:** This type of analysis distinguishes between the opinions and the facts found in the data. An internet remark about changing a battery incites hostility towards customers. Figure 2 provides an example of how customer service might interact with the system to resolve issues.



**Figure 2 Sentiment types Source Expressanalytics.com**

Among sentiment analysis's benefits are

**Sorting Data at Scale:** The system could be more creative in how it displays the quantity of tweets, customer support conversations, or survey replies. In the current setting, human intervention is used in the collection and processing of all corporate data. Additionally, sentiment analysis processes a large amount of unstructured data in order to provide the system with the most effective and economical analysis reports.

**Real-time analysis:** A public relations (PR) crisis that results in unhealthy business practices is one of the rapidly growing crises in real time. Sentiment analysis is used to identify major affairs. In addition, sentiment analysis models can help quickly identify unhappy customers and take appropriate action to get rid of them.

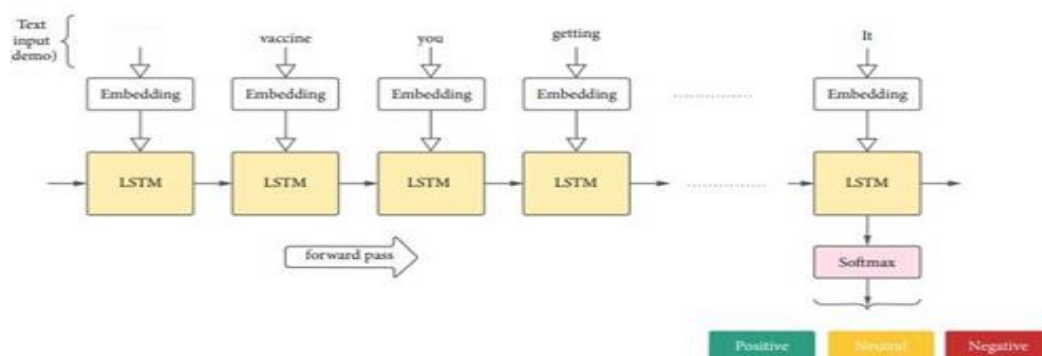
**Consistent standards:** Approximately 60–65 percent of respondents concur when asked to assess the text's sentiment. Text sentiment tagging can be influenced by human experiences, ideas, and beliefs. By using a centralised sentiment analysis system, overall performance, including accuracy and insightful analysis, can be enhanced while maintaining the same conditions for the company's data.

Model preparation mistakes are often the root cause of sentiment analysis difficulties. Frameworks struggle with objectivity, or unbiased feeling statements, which are often misunderstood. For instance, a review stating, "The item looked blue," written by a customer after receiving an unsatisfactory variety item would be categorised as neutral as opposed to negative. It can be challenging to interpret sentiment when frames are unable to capture the distinct tone or context. It is currently difficult to sort responses to study questions and questionnaires that include the words "anything" or "any" when the setting isn't specified. They could receive a positive or negative label based on the question. In addition, it can be challenging to teach incongruity and mocking, and many mislabeled emotions.

Statements on their own could be difficult. Positive and negative feedback will almost always be included in reviews; this can be handled by going over each statement individually. However, people are more likely to combine different points of view in a single statement in informal communication, which makes it harder for a computer to understand. Furthermore, machine learning techniques are taught to train sentiment analysis within

frameworks that are programmed. Logic-based sentiment analysis techniques include perfectly acceptable sentiment polarity, feeling recognition, viewpoint sentiment analysis, and anticipation assessment.

### SYSTEM ARCHITECTURE



**Figure 3. LSTM for sentiment class mapping**

Finding a sentence's dependency parse in order to comprehend the connection between the "head" words is known as dependency parsing. Dependency parsing makes any sentence easier to understand by giving it a syntactic structure. An analysis of a sentence's syntactic and semantic structure can be done using these kinds of syntactic structures. In other words, the parsing tree can verify the sentence's semantic structure in addition to its grammar. The most common syntactic structure is the parse tree, which can be produced by parsing algorithms such as the Earley algorithm (Figure 3 illustrates LSTM) or the Chart parsing algorithm. The dynamic programming in each of these algorithms allows it to overcome ambiguity issues. Assigning the syntactic structure can get very complicated because a given sentence can have multiple dependency parses. Uncertainties, or multiple parse trees, must be cleared up for a sentence to have a clear syntactic structure. Syntactic disambiguation is the process of selecting the correct parse from a set of multiple parses where each parse has certain probabilities.

#### Algorithm

**Step 1:** Splitting the trained dataset using ``process_row`` and ``tokenizer`` methods.

**Step 2:** Obtaining the data in a multidimensional array.

**Step 3:** For all indices in the data, perform sentiment analysis on each index as follows:

- Sentences: Text split using periods (".").
- Tokenizer: AutoTokenizer from pretrained ("bert-base-uncased") follows the algorithm BERT (Bidirectional Encoder Representations from Transformers).

**Step 4:** Processing each new row and storing the overall results using multi-threading with ``ThreadPoolExecutor()`` as an executor.

- Overall\_results: List comprehension using ``executor.map(self.process_row, df.itertuples(index=False))`` follows the proposed algorithm for grouping final results.

**Step 5:** Save the overall calculated sentiment analysis results and label them based on categorization such as fully positive, fully negative, and partial.

**Step 6:** Assign the label for each row as 1, 1, 0.00... based on the categorization done above.

### PERFORMANCE ANALYSIS METRICS

A range of metrics are employed to assess the machine learning algorithms' performance results. Performance metrics must be used to validate regression algorithms. The comprehension of data balance is provided by the metric formulation. The performance metric must be consistent across all data, and the same

formula for the performance metric is used to compare one or more analysis models in further detail. The attributes of the outcome will differ based on the chosen metric. Confusion Matrix: Information from several classes is available in the input data. The correctly classified values and non-correct values form the basis of the classification accuracy. The True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) values are used to gauge the relationship between the right and incorrect classification of the data.

When the maximum score of 1 is correlated with both the actual and predicted data patterns, it is considered a

**True Positive (TP):** True Positives is defined as the case when both actual data pattern and predicted data pattern correlates to the maximum score of 1.

**True Negative (TN):** A true negative describes a situation in which the expected pattern closely resembles the observed pattern.

**False Positives (FP):** FP is the term for when true data is predicted to be positive but only produces one real case instead of zero.

**False Negatives (FN):** A false negative occurs when a pattern in the data causes the prediction to go negative, allowing the prediction system to reflect a positive result in this case, with actual = 0 and predicted class = 1.

**Classification Accuracy:** One of the key metrics used to assess the effectiveness of classification algorithms is accuracy. The ratio of all predictions made during the process to the total number of values that were correctly classified is known as accuracy. The confusion matrix is used to assess the expression below.

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

### **III. RESULTS AND DISCUSSIONS**

#### **Experimental results**

##### **Dataset Description**

We have tweets and amazon to form training and testing dataset. These datasets are cleaned for the further processing.

##### **Experimental setup**

The RNN based LSTM model is built using Keras and Tensorflow.

Parameters	Value
Batch-size	10
Epochs	128
Hidden layer-activation function	Relu
Output layer-activation function	Sigmoid
Optimizer learning-rate	Adam 0.001
Dropout	0.5
Loss-function	Binary cross-entropy

**Table 1 LSTM Parameter**

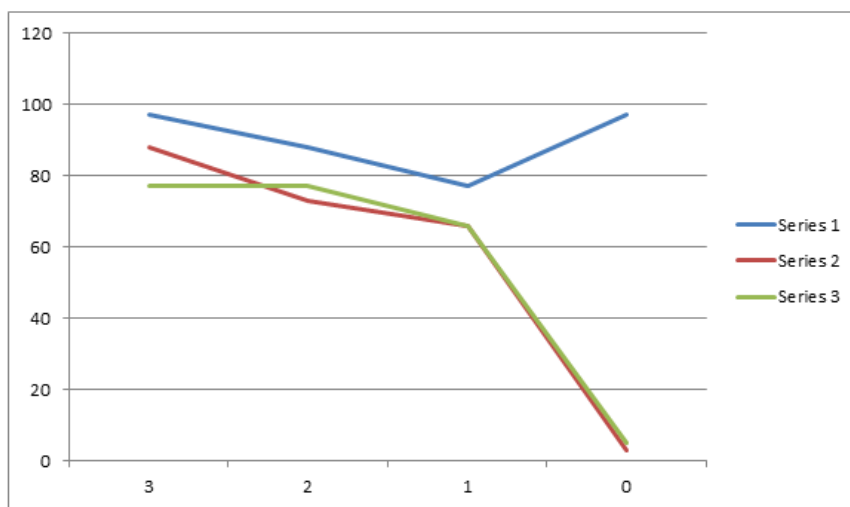
Table 1 represents the network configuration parameters of the RNN-LSTM model.

**Performance evaluation**

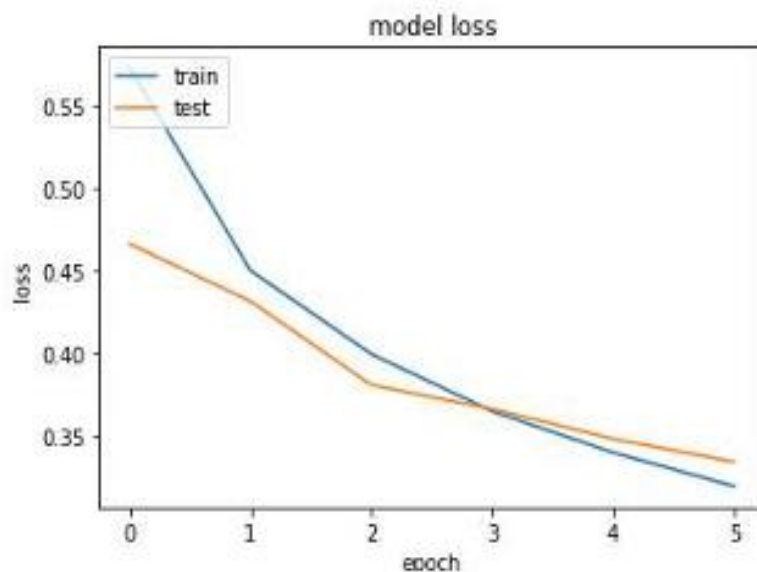
In this subsection, we will evaluate the effectiveness of the used methodology and other relevant strategies. The collection of positive feelings derived from Twitter data.

Original Text	Sentiment Label	Overall Confidence
user when a father is dysfunctional and is so selfish he drags his kids into his dysfunction run	NEGATIVE	1.0
user user thanks for lyft credit i cant use cause they dont offer wheelchair vans in pdx disapointed getthanked	NEGATIVE	1.0
bihday your majesty	POSITIVE	1.0

**Figure 4 Sentiment Analysis**



**Figure 5 epocs**



**Figure 6. Model Performance**

The LSTM model classifies feelings with 97% accuracy. Figures 4,5,6 show the accuracy and confusion matrix loss of our model in each iteration throughout 5 epochs.

The LSTM technique outperforms the Convolutional Neural Network (CNN) and the simple neural network. Table 2 compares the performance of the algorithms for tweets and reviews.

Algorithm Name	Training Accuracy	Testing Accuracy
LSTM	97	97.01
CNN	92	82
NN	84	83

**Table 2. Performance Matrix-I**

Table 3 indicates the performance of the three different methods based on four parameters.

Methods	Accuracy	Precision	Recall	F-Measure
LSTM	97	91	90	90
CNN	92	77	76	76
NN	84	76	75	74

**Table 3. Performance matrix-II**

#### IV. CONCLUSIONS AND RECOMMENDATIONS

The proposed research addressed the requirement for RNN and LSTM to improve Sentiment classification performance. The report also addresses the various AI-based techniques and their performance. We evaluated the performance of the LSTM-RNN model. We examined and reviewed tweets for our research, categorising them as positive or negative sentiments. The proposed strategy has a 97% accuracy level, which is the highest among all ways employed in the existing system. To conduct the experiment, we obtained a Twitter dataset from the Kaggle source. Although the proposed LSTM-based solution is the best, it requires real-time data analysis and emotion classification. To address scalability concerns, a big data platform can be used to analyse the huge number of tweets.

#### REFERENCES

- [1]. Hochreiter, Sepp & Schmidhuber and Jürgen Long Short-term Memory 1997 Neural Computation 9 1735-80.
- [2]. Bengio Y, Simard P and Frasconi P 1994 Learning long-term dependencies with gradient descent is difficult IEEE Transactions on Neural Networks 5 2 157- 66.



- [3]. Mike Schuster and Kuldip K 1997 Bidirectional recurrent neural networks IEEE Transactions on Signal Processing 45 11 2673–2681.
- [4]. Patel P, Patel D and Naik C 2021 Sentiment Analysis on Movie Review Using Deep Learning RNN Method.
- [5]. Satapathy S, Zhang Y D, Bhateja V, Majhi R 2021 Intelligent Data Engineering and Analytics J Advances in Intelligent Systems and Computing 2 1177.
- [6]. Anand M, Sachin S 2017 Proc. 5th Int. Conf. on Mining Intelligence and Knowledge Exploration: MIKE 2017 (Hyderabad, India) Sentiment Analysis of Tweets in Malayalam using Long Short-Term Memory Units and Convolutional Neural Nets.
- [7]. Deriu, Jan & Cieliebak, Mark 2016 Sentiment Analysis using Convolutional Neural Networks with Multi-Task Training and Distant Supervision on Italian Tweets.
- [8]. Fenna Miedema and Sandjai Bhulai 2018 Sentiment Analysis with Long Short-Term Memory network Vrije Universiteit Amsterdam.
- [9]. Zainuddin, Nurulhuda & Selamat, Ali 2014 Proc. Of 1st Int. Conf. on Computer, Communications, and Control Technology Sentiment analysis using Support Vector Machine pp 333-37.
- [10]. Mathieu Cliché, Proc. of the 11th International Workshop on Semantic Evaluations (SemEval 2017) pp 573–80.
- [11]. Go A, Bhayani R and Huang L 2009 Twitter sentiment classification using distant supervision CS224N Project Report Stanford 1
- [12]. Tomas Mikolov, Kai Chen, Greg Corrado and Jeffrey Dean 2013 Efficient estimation of word representations in vector space arXiv preprint arXiv: 1301.3781.
- [13]. Danqi Chen and Christopher D Manning 2014 Proceedings of the Empirical Methods in Natural Language Processing (EMNLP) A fast and accurate dependency parser using neural networks pp 740–50.
- [14]. Jeffrey Pennington, Richard Socher and Christopher Manning 2014 Proceedings of the Empirical Methods in Natural Language Processing (EMNLP) Glove: Global vectors for word representation pp 1532–43
- [15]. D. Mali, M. Abhyankar, P. Bhavarathi, K. Gaidhar, M. Bangare International Journal of Management and Applied Science, ISSN: 2394- 7926 Volume-2, Issue-1, Jan.-2016
- [16]. Duc-Hong Pham, Anh-Cuong Le, 2018 International Journal of Approximate Reasoning (2018)
- [17]. Duc-Hong Pham , Anh-Cuong Le, 2018, Data & Knowledge Engineering(2018)
- [18]. Anvar Shathik J. & Krishna Prasad K. ,2020, International Journal of Applied Engineering and Management Letters (IJAEML), ISSN: 2581-7000, Vol. 4, No. 2, August 2020.
- [19]. Wenhao Zhang, Hua Xu , Wei Wan ,2012 Expert Systems with Applications 39 (2012)
- [20]. Fu Xianghua , Liu Guo, Guo Yanyan, Wang Zhiqiang. ,2013, Knowledge-Based Systems 37 (2013).
- [21]. Raktim Kumar Dey, Debabrata Sarddar, Indranil Sarkar, Rajesh Bose, Sandip Roy,2020, International Journal Of Scientific & Technology Research Volume 9, Issue 05, May 2020 ISSN 2277-8616.
- [22]. W.A. Awad and S.M. Elseuofi, 2011, International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 1, Feb 2011.
- [23]. Anas Collomb, A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation, IEEE TRANSACTIONS ON MULTIMEDIA, 2016
- [24]. Shreya Banker and Rupal Patel, A BRIEF REVIEW OF SENTIMENT ANALYSIS METHODS International Journal of Information Sciences and Techniques (IJIST) Vol.6, No.1/2, March 2016
- [25]. Bruno Ohana , Brendan Tierney, Sentiment Classification of Reviews Using SentiWordNet 2016
- [26]. Xiaojiang Lei, Xueming Qian, Rating Prediction Based on Social Sentiment From Textual Reviews, IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 18, NO. 9, SEPTEMBER 2016
- [27]. Evangelos Psomakelis, Konstantinos Tserpes, COMPARING METHODS FOR TWITTER SENTIMENT ANALYSIS Conference: International Conference on Knowledge Discovery, 2017.