IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

SENTIMENTAL ANALYSIS USING TREE ATTENTION ALGORITHM

GUIDE :
Dr.P.EDWIN DHAS,ME,Ph.D,

Students(UG):
1 ,A.Arockya Joseph Raj
2 ,M.Prakashraj
3 , P.Ebanesh
4 ,K.Augustin

,Department of computer science and engineering, Jayaraj Annapackiam CSI College of Engineering, Nazareth.

Abstract

The goal of this project is to develop a system for computational evaluation of user feedback to detect emotional tone using advanced language processing techniques, specifically employing the Tree Attention Algorithm. The system will analyze text data and automatically determine the emotional tone or attitude expressed by the user, such as positive, negative, or neutral. To accomplish this, we will utilize a combination of machine learning algorithms and deep learning models, including the Tree Attention Algorithm, which has been shown to outperform other state-of-the-art techniques in sentiment analysis tasks. The system will also incorporate pre-processing techniques such as tokenization, stemming, and stop-word removal to enhance the accuracy of emotion detection. We will evaluate the system using a dataset of user feedback, including reviews, comments, and social media posts, and compare its performance with existing sentiment analysis tools. The system's accuracy and efficiency in detecting emotional tone will be assessed using standard evaluation metrics such as precision, recall, and F1 score. The proposed system has numerous applications, including brand reputation management, customer feedback analysis, and market research. The ability to automatically detect emotional tone in user feedback can help businesses and organizations better understand their customers' needs and preferences and make data-driven decisions to improve their products and services.

Introduction

In today's digital world, user feedback is a valuable resource for businesses and organizations to better understand their customers' needs and preferences. However, analyzing large volumes of text data manually can be time-consuming and impractical, especially when trying to determine the emotional tone or sentiment expressed by the user. This is where computational evaluation using advanced language processing techniques comes in. The goal of this project is to develop a system that can automatically detect emotional tone in user feedback using advanced language processing techniques and the Tree Attention Algorithm. The system will be able to analyze text data and determine whether the user's feedback is positive, negative, or neutral, which can provide valuable insights into customer satisfaction and brand reputation. We will use a combination of machine learning algorithms and deep learning models to develop the system, which will incorporate pre-processing techniques such as tokenization, stemming, and stop-word removal to improve its accuracy. The Tree Attention Algorithm will be used as the primary deep learning model, as it has been shown to outperform other state-of-the-art techniques in sentiment analysis tasks. The system will be evaluated using a dataset of user feedback, including reviews, comments, and social media posts. The accuracy and efficiency of emotional tone detection will be assessed using standard evaluation

metrics such as precision, recall, and F1 score. The system's performance will also be compared to existing sentiment analysis tools to determine its effectiveness. Overall, this project aims to provide a valuable tool for businesses and organizations to analyze user feedback more efficiently and accurately, enabling them to make data-driven decisions to improve their products and services and better understand their customers' needs and preferences. User feedback is a critical aspect of customer satisfaction and brand reputation management. However, analyzing large volumes of text data manually can be time-consuming and impractical, especially when trying to determine the emotional tone or sentiment expressed by the user. This is where computational evaluation using advanced language processing techniques comes in. The ability to automatically analyze user feedback and determine the emotional tone can provide valuable insights into customer satisfaction and brand reputation. The aim of this project is to develop a system for computational evaluation of user feedback to detect emotional tone using advanced language processing techniques, specifically employing the Tree Attention Algorithm. Sentiment analysis or opinion mining is a subfield of natural language processing that involves identifying and extracting subjective information from text data. The goal of sentiment analysis is to automatically determine the emotional tone or attitude expressed in a piece of text, such as a review or social media post. Sentiment analysis has many practical applications, such as in social media monitoring, brand reputation management, customer feedback analysis, and market research. There are several approaches to sentiment analysis, including rule-based methods, machine learning algorithms, and deep learning models. Rule-based methods rely on manually crafted rules to identify sentiment indicators, while machine learning algorithms use statistical techniques to learn patterns in labeled training data. Deep learning models, such as convolutional neural networks and recurrent neural networks, can learn to extract features and representations from raw text data, allowing them to achieve state-of-the-art performance on sentiment analysis tasks.

Related Works

There has been an exponential growth in the number of complex documents, texts, and user-generated content, mainly through social media, blogs and review websites, etc., that require a deeper understanding of deep learning methods to be able to accurately classify texts. This research work approaches problems of fine-grained sentiment analysis. We explore how the sentiment of the user-generated text can be identified using Deep Learning. This paper contributes sentiment analysis, which aims to extract opinions from the text. E-commerce, social media, and other many online user-generated contents provide a powerful way for collecting people's thoughts and their feelings towards users or customers. Powerful sentiment analysis helps many businesses with regard to sentiment, so they will understand users' perceptions of the product. It's also very helpful for other customers to get reviews of products and get to know about services they offer. Lots of websites provide users with reviews but it's very difficult to make some decisions from them. Deep Learning sentiment helps to identify the pros and cons of a product. Nowadays Automated sentiment analysis systems are tools for decision-making for many parties. The deep learning algorithms are able to handle complex sequential data and determine non-linear relationships within data. Recurrent Neural Networks (RNN) are widely used in natural language processing (NLP) because they are very suitable to process variable-length text. Doing Fine-grained sentiment analysis is a challenging task. We use a modified Recurrent Neural Network (Tree-LSTM) for fine-grained analysis. In this work, we provide a brief overview of deep learning classification algorithms and also conclude that Tree-LSTM gives state-of-the-art accuracy for fine-grained sentiment analysis. [1]

In this study, a method is proposed to represent the corresponding viewpoints and features in the form of graphs for observing the sentiment of the subscribers based on product review. The aim is to observe and empathize the situations based on real time with sentimental analysis using deep learning and machine learning. Nowadays machine learning and deep learning is applied for the solutions of the problems in computer vision, voice recognition, image processing and convolutional neural network. In this work, a convolutional neural network model has been proposed which attains a better outcome than other methods. [2]

The corporate environment has gotten more competitive in recent years. Customer pleasure has shifted to the forefront of company strategy. Businesses spend a lot of money and human resources on different approaches to understand and satisfy their consumers' needs. However, many firms fail to attain customer satisfaction as a result of poor manual analysis of a wide range of consumer demands. Because of this, they are shedding consumer commitment and incurring additional marketing costs. Sentiment Analysis may be used to solve the challenges. It combines Natural Language Processing (NLP) and Machine Learning techniques (ML). Sentiment Analysis is a technique for extracting information from public opinion on a wide range of subjects, goods, and services. We can accomplish that using any publicly available data online. On a look for a powerful strategy for Sentiment Analysis on a big, unbalanced, and multi-class dataset, we used two NLP approaches (Bag-of-Words and TF-IDF) and multiple ML classification algorithms (Support Vector Machine, Logistic Regression, Multinomial Naive Bayes, Random Forest). Sentimental analysis, or SA, is a continuing branch of research that examines customer feelings when purchasing things. This research report looks at a comprehensive mobile phone reviews study. Customer segmentation is an important part of every company's marketing strategy. In this machine learning project, we'll utilize nltk to cluster and manage unlabeled datasets. The findings reveal that the method works, with high accuracy in both mobile classification and user segmentation. [3]

Social networking platforms have become an essential means for communicating feelings to the entire world due to rapid expansion in the Internet era. Several people use textual content, pictures, audio, and video to express their feelings or viewpoints. Text communication via Web-based networking media, on the other hand, is somewhat overwhelming. Every second, a massive amount of unstructured data is generated on the Internet due to social media platforms. The data must be processed as rapidly as generated to comprehend human psychology, and it can be accomplished using sentiment analysis, which recognizes polarity in texts. It assesses whether the author has a negative, positive, or neutral attitude toward an item, administration, individual, or location. In some applications, sentiment analysis is insufficient and hence requires emotion detection, which determines an individual's emotional/mental state precisely. This review paper provides understanding into levels of sentiment analysis, various emotion models, and the process of sentiment analysis and emotion detection from text. Finally, this paper discusses the challenges faced during sentiment and emotion analysis. [4]

Sentiment analysis (SA) is becoming more and more irreplaceable when we talk about recommendation or analyze guest's preference. Sentiment classification gives prediction ratings or classifying the emotion into two or more poles. We are now looking forward to perfect the precision of the prediction of models. Nowadays, with the development of neural network, deep learning has been broadly used in sentiment analysis. In this paper, we survey large number of papers concerning sentiment classification with deep learning. We propose a new venue to make a division of all the models into 6 kinds. Moreover, we make a comparison of these results of models and draw a conclusion. [5]

System Analysis

Analysis is the first crucial step, Analysis is defining the boundaries of the system that will be followed by design and implementation. This section gives the detailed study of the various operations performed by a system and their relationships within and outside of the system.

Existing Algorithms

Existing Algorithms namely: support vector machine (SVM), Recurrent Neural Networks (RNN), Naïve Bayes (NB) and Long Short Term Memory (LSTM).

SVM

SVM classification algorithm to classify customers according to their buying behaviour. Classification is done by considering how the customer spends their valuable time, day in buying decisions.

RNN

Recurrent Neural Networks (RNNs) are a type of neural network commonly used in Natural Language Processing (NLP) tasks such as sentiment analysis. RNNs are particularly suited to processing sequential data, such as text, because they can capture the temporal dependencies between words in a sentence or document. In sentiment analysis, RNNs are often used to classify text as positive, negative, or neutral. The input to the RNN is a sequence of words, which are represented as word embeddings. The RNN processes the sequence of word embeddings one word at a time, and at each time step, it updates its hidden state based on the previous hidden state and the current input. The final hidden state is then used to make the sentiment classification.

Naive Bayes

Naive Bayes classification algorithms based on Bayes' theorem which is powerful to the predicted variables. Naïve Bayes algorithm is to is classified the group of data items to efficient and, correct, fast. It is more accept in different group of data prediction analysis. When we assume of non-dependence data variables is handle, a Naive Bayes algorithm to perform the good compare to other model like regression analysis. It is good performing give to the different input data compare the numeric value of variable in data, for numeric value of distribution is predicted.

LSTM

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture that is commonly used in sentiment analysis tasks. LSTMs were developed to overcome the vanishing gradient problem that is often encountered when training traditional RNNs on long sequences of data. In sentiment analysis, LSTMs are used to model the sequence of words in a sentence or document and to predict the sentiment of the text. LSTMs can capture long-term dependencies in the text by selectively retaining or forgetting information from previous time steps, based on the current input.

Disadvantages

- Machine Learning needs enough time to let the algorithms learn and develop enough to fulfil their purpose with a considerable amount of accuracy and relevancy.
- Machine Learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality.
- Not able to accurately interpret results generated by the algorithms.
- Highly susceptible to errors.

Proposed System

This project proposed a Sentimental analysis system using a Tree Attention Algorithm has been developed for Analysing whether the content in the videos are positive, negative or neutral.

TREE ATTENTION

These are like an upgrade over LSTMs. In bidirectional LSTMs, each training sequence is presented forward and backward so as to separate recurrent nets. Both sequences are connected to the same output layer. Bidirectional LSTMs have complete information about every point in a given sequence, everything before and after it.

The human brain uses its senses to pick up information from words, sounds, or from whole sentences that might, at first, make no sense but mean something in a future context. Conventional recurrent neural networks are only capable of using the previous context to get information. Whereas, in bidirectional LSTMs, the information is obtained by processing the data in both directions within two hidden layers, pushed toward the same output layer. This helps bidirectional LSTMs access long-range context in both directions.

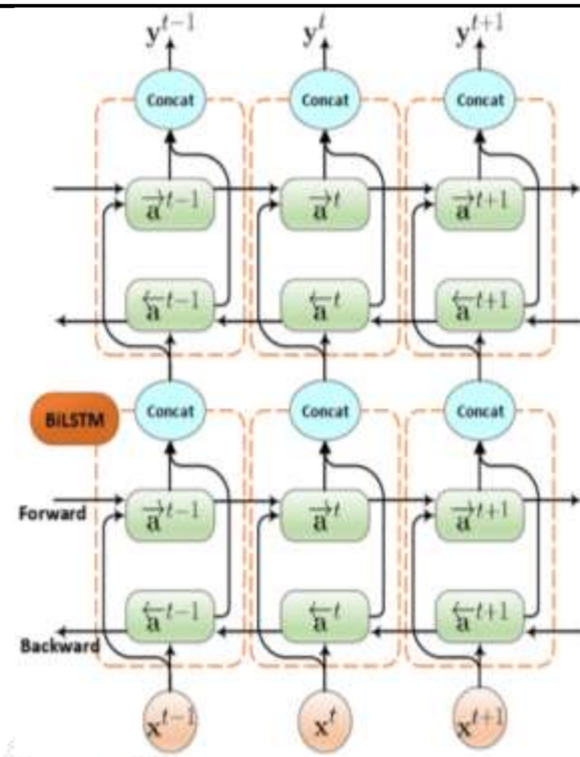


Figure .3.1. Bi-LSTM

CNN

CNN is a kind of feed-forward neural network used in deep learning, which was originally used in computer vision and included a convolutional layer to create the local features and a pooling layer for summarizing the representative features. Convolution layers in the artificial neural network play the role of a feature extractor that extracts the local features. This means that CNN establishes the specific local communication signals using a local connection pattern between neurons in the adjacent layer. Such a feature is useful for classifying in NLP, as it is expected that strong local clues should be found for the class, but these clues may appear in different places at the input. The convolutional and pooling layers allow CNNs to find local indicators, regardless of their location.

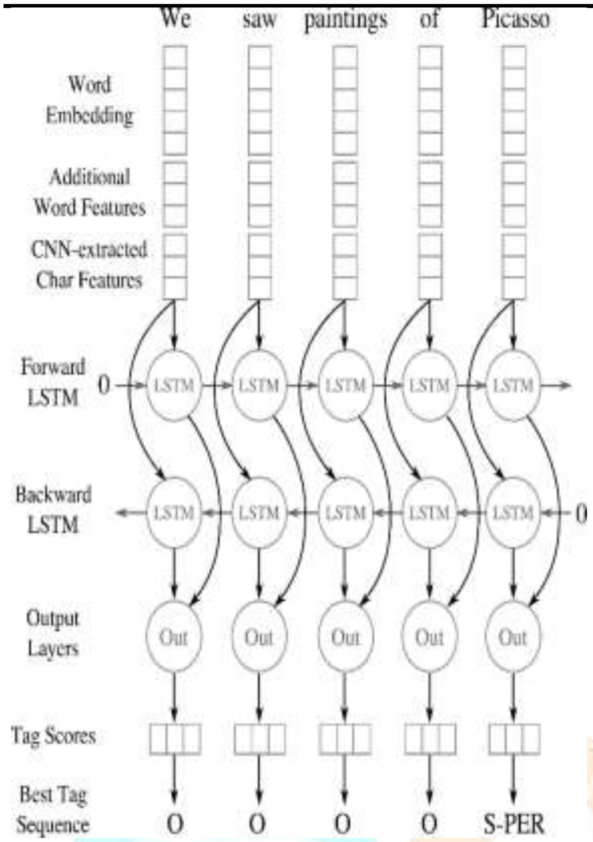


Figure.3.2. CNN-BiLSTM

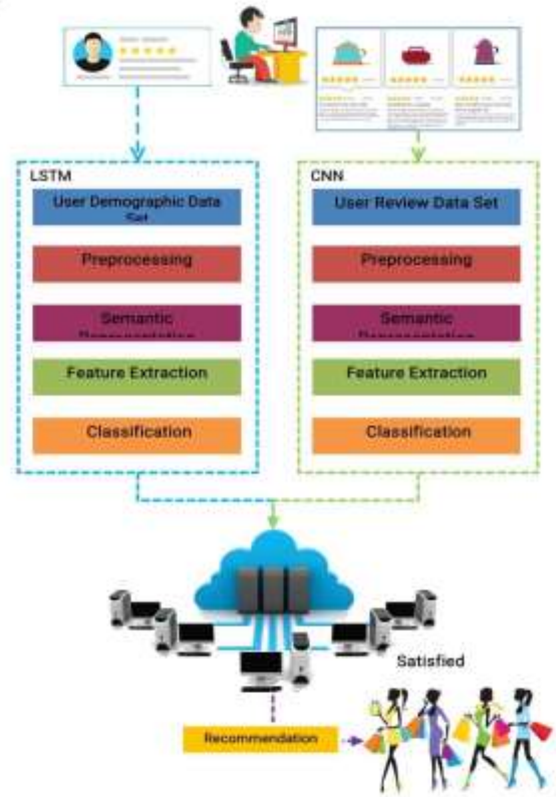
Advantages

- Recommendations had to be revenue driven, maximizing profits of the service, while providing high quality services to a customer.
- customer behaviour prediction without any hand-crafted features.
- Recommendation based on user behaviour
- Accurate predictions

System Architecture

1.Consumer Behaviour Analysis API

In this project we developed the web application for consumer behavior analytics on amazon consumer demographic and product review data.



2. Data Set Annotation

In Training Phase “We used product review data from Kaggle. The dataset in consisted of two labels, positive and negative, while was composed of three labels of positive, neutral, and negative. Furthermore, the dataset in was composed of five labels of positive, somewhat positive, neutral, somewhat negative, and negative.”

In Testing Phase “In this project, user login and search, purchase, and give reviews about the product and also search the product based on the search the user gets best recommended products.

3.Pre-processing

In the first step, the data should be pre-processed to reduce the implementation time and improve the results. In this module Preprocessing was carried out to modify the text data appropriately in the experiment. We used decapitalization and did not mark the start and end of the sentences. The system deleted #, two or more spaces, tabs, Retweets (RT), and stop words. The application also changed the text that represented the url that began with “http” to [URL] and the text that represented the account ID that began with “@” to [NAME].

In addition, then changed digits to [NUM], and special characters to [SPE]. The system changed "can't" and "isn't" to "cannot" and "is not", respectively, since "not" is important in sentiment analysis.

4.Feature Extraction

Convolutional Neural Network (CNN) is a special kind of deep neural network model. We design a double-layer parallel convolutional neural network to extract and represent the short text features.

Convolution Layer For Feature Extraction

The purpose of convolution layer is to extract semantic features of the sentence, each convolution kernel corresponds to a certain part of feature and the feature mappings can be obtained after convolution operation.

K-Max Pooling Layer For Feature Dimension Reduction

Features extracted by the convolution layer are transmitted to the pooling layer which will further aggregate and simplify the feature representation. K-Max pooling is adopted to select the top-K value of each filter to represent the semantic information. The larger the feature value, the greater the emotional strength.

5.Classification

The feature matrix extracted by CNN model or the semantic matrix extracted by Capsule is input into a dropout layer to prevent the over-fitting problem. During the training process, some neurons which are selected randomly in the hidden layer do not work, but they are still retained for the next input sample. The other neurons participate in the process of computation and connection. The vector matrix is input into a full connection layer for dimension reduction. Finally, the probability distribution of the sentiment category is computed by softmax activation function y D soft max(x).Then the classified results is stored to database as a csv file format.

CNN consists of:

Convolutional layer -> extracts different features of the input -> more layers extract complex features from the last feature.

Pooling layer -> combines outputs of neuron clusters at one layer into a single neuron in the next layer.

Fully-connected layers -> combines all local features into global features to calculate final result.

CNN can automatically extract and learn features. CNN uses highly dimensional data with minimal processing.

Results and Discussion

According to the combination of the real category of the sample and the predicted category of the model in the paper, the results can be divided into four types: true positive (TP), false-positive (FP), true negative (TN), and false-negative (FN). The larger the value on the main diagonal is, the smaller the value on the sub-diagonal is, the better the model. After digitizing the confusion matrix, it is Precision, Recall, and F1-score. For the prediction problem of e-commerce consumers repurchase behavior, the accuracy and AUC value are added to evaluate the training and test results of the model. The specific formula is as follows:

Precision, the ratio of the sum of true positive (TP) to true positive (TP) and false-positive cases (FP):

$$Precision = \frac{TP}{TP + FP}$$

Recall the ratio of the sum of true positive (TP) to true positive (TP) and false-negative cases (FN):

$$Recall = \frac{TP}{TP + FN}$$

F1-score is the overall evaluation and weighted average based on precision and recall. "1" represents the best score of "F1", while "0" represents the worst score. The formula is as follows:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Accuracy, the ratio of the number of correctly classified samples to the total number of samples. The standard cannot reflect the potential distribution of response value, nor can it output the classifier's error types. Nevertheless, it is easier to understand. The formula is as follows:

$$Accuracy = \frac{TP + tn}{(TP + FP) + (TN + FN)}$$

AUC (Area under the curve) is an evaluation indicator used for the binary classification model, indicating that a positive sample and a negative sample are randomly selected. The classifier can correctly give the probability that the score of the positive sample is higher than the negative sample, the area under the ROC curve. The larger the area, the better the model effect is, and the formula is as follows:

$$AUC = \frac{\int_0^1 TPdFP}{(TP + FN)(TN + FP)}$$

Conclusion

This project aims to develop a computational approach for evaluating user feedback to detect the emotional tone using advanced language processing techniques and a tree attention algorithm. The system will be designed to automatically detect the sentiment of text data from various sources, such as customer reviews, social media posts, and feedback forms.

The system design involves data preprocessing, feature extraction, and sentiment classification using the Tree Attention algorithm. The system will be trained on a large dataset of annotated text data to ensure accurate and reliable sentiment analysis results. The project also includes an evaluation of the system's performance using metrics such as precision, recall, and F1-score.

The proposed system has the potential to provide valuable insights into user sentiment and emotional tone, which can be used to improve customer satisfaction and enhance business operations. The system can be applied to various domains, including healthcare, finance, politics, and social media.

Future Enhancement

Expanding the project to include analysis of sentiment in multiple languages. This would require training the model on a larger and more diverse dataset of multilingual text. Improving the model's ability to recognize and account for contextual nuances in sentiment analysis. This could involve incorporating contextual information such as the topic of the text, the speaker's demographic information, or the tone of the text. Developing domain-specific sentiment analysis models that are trained on a particular domain, such as finance, healthcare, or politics. This would allow for more accurate sentiment analysis in specific industries or areas of interest. Expanding the model to include the detection of specific emotions such as anger, joy, or sadness. This would require a more fine-grained analysis of the sentiment in the text. Developing a real-time sentiment analysis system that can analyze sentiment in real-time as data is collected, such as from social media streams or customer feedback channels. Improving the interpretability and explain ability of the sentiment analysis model. This would involve developing tools and methods to explain how the model arrived at a particular sentiment classification or prediction.

References

[1] Sentiment Analysis By Using Modified RNN And A Tree LSTM, International Conference on Computing Communication and Power Technology (IC3P), January 2022[Online].Available: https://ieeexplore.ieee.org/document/979348

[2] Sentiment Analysis to Review Products based on Machine Learning,4th International Conference on Inventive Research in Computing Application(ICIRCA),September2022.[Online]. Available: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00308-7

[3] Product Recommendation Using Sentiment Analysis, International Conference on Engineering and Emerging Technologies (ICEET), October 2022, [Online].Available: https://ieeexplore.ieee.org/document/10007234

[4] A review on sentiment analysis and emotion detection from text, August 2021, [Online]. Available: https://link.springer.com/article/10.1007/s13278-021-00776-6

[5] A Survey of Sentiment Analysis Based on Product Review ,January 2021, [Online].Available: https://ieeexplore.ieee.org/document/9463281

[6] Comprehensive Study on Sentiment Analysis: Types, Approaches, Recent Applications, Tools and APIs, July 2020, [Online].Available: https://ieeexplore.ieee.org/document/9213209.

[7] R. Sussman and R. Gifford, "Be the change you want to see modeling food composting in public places", Environment and Behavior, vol. 45, no. 3, pp. 323-343, 2013, [Online]Available: https://journals.sagepub.com/doi/10.1177/0013916511431274.

[8] Research on sentiment analysis technology and polarity computation of sentiment words, December 2010, [Online].Available: https://ieeexplore.ieee.org/abstract/document/5687438

[9] H. X. Shi, G. D. Zhou and P. D. Qian, "An attribute-based sentiment analysis system", Information Technology Journal, vol. 9, no. 8, pp. 1607-1614, 2010.

[10] S. Baccianella, A. Esuli and F. Sebastiani, "Multi-facet rating of product reviews", Toulouse France ECIR'2009, pp. 461-472, 2009.

[11] Q. Miao, Q. Li and R. Dai, "AMAZING: A sentiment mining and retrieval system", Expert Systems with Applications, vol. 36, no. 3, pp. 7192-7198, 2009.

[12] R. McDonald, K. Hannan, T. Neylon, M. Wells and J. Reynar, "Structured models for fine-to-coarse sentiment analysis", Prague Czech Republic ACL'2007, pp. 432-439, 2007.

[13] J. J. Liu, Y. B. Cao, C.Y. Lin, Y. L. Huang and M. Zhou, "Low-Quality product review detection in opinion summarization", Prague Czech Republic EMNLP'2007, pp. 343-350, 2007.

[14] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", LREc, vol. 10, no. 2010, 2010.

[15] P. Nakov, A. Ritter, S. Rosenthal, Sebastiani and V. Stoyanov, "SemEval-2016 Task 4: Sentiment Analysis in Twitter", ACLWEB, 2016, [online] Available: http://bvicam.ac.in/indiacom/Downloads.asp

[16] Wehrmann Joonatas, Willian Becker, Henry EL Cagnini and Rodrigo C. Barros, "A character-based convolutional neural network for language-agnostic Twitter Sentiment Analysis", *IEEE*, 2017.

[17] B. Liu and L. Zhang, "A Survey of Opinion Mining and Sentiment Analysis", *Springer Link*, 2012.

[18] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications", *ScienceDirect*, 2012.

[19] Adyan Marendra Ramadhani and Hong Soon Goo, "Twitter Sentiment Analysis using Deep Learning Methods" in , Department of Management Information Systems Dong-A University Busan South Korea, 2017.

[20] Rincy Jose and Varghese S Chooralil, Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Classifier Ensemble Approach, Ernakulam, India:Department of Computer Science and Engineering Rajagiri School of Engineering and technology, 2016.