Sentiment analysis for Product Reviews

Anurag Bairagi

School of Computing Science & Engineering, Galgotias University

Greater Noida, India

Sherya Singh

School of Computing Science & Engineering, Galgotias University

Greater Noida, India

Abstract:

Sentiment analysis is an effective method for identifying text data and extracting the sentiment component. Every day, customers' reviews, opinions, suggestions, and tweets generate a high amount of unstructured data on shopping websites. Retailers can use aspect level analysis of this data to gain a better knowledge of their customers' expectations and then modify their policies accordingly. A innovative approach based on aspect level sentiment detection, which focuses on the item's features, is provided in this research. The work was implemented and validated on Amazon customer reviews (crawled data), where each review's aspect phrases were determined initially. The technique pre-processes the dataset to extract valuable information, such as stemming, tokenization, casing, and stop-word removal. After preprocessing the data and cleaning it using various ways, some useful features are chosen and sentiment analysis is performed to generate a sentiment polarity. On the preprocessed data, various learning methods such as Logistic Regression classifiers, Linear Support Vector Machine and Naïve Bayes are applied, and a comparison analysis is performed to determine the best classifier fit for the reviews data through detailed analysis and generation of the Receiver operator characteristics curve, as well as a comparison analysis through AUC values of different classifiers and then assigns a rank to its classification in negativity or positivity.

Keywords: Sentiment analysis, pool based active learning, feature extraction, text classification, machine learning.

Introduction:

Sentiment is a person's experience or opinion on a certain product, topic, or event. Hence sentiment analysis, also known as opinion mining, examines people's opinion of specific entities. Today's large retailers are rapidly growing their locations. They also collect user reviews for the items they sell. Users can write product reviews on the retailer's website or share their information to other social media platforms and forums. As a result, a big pool of product opinions is easy to access on the internet. If this pool is carefully evaluated, it can provide a wealth of information on the product's quality factors. However, since the data in every evaluation is not in a logical order, extracting the important details from the data available is a main challenge. Sentiment analysis, which is based on text mining and natural language processing (NLP) technology, delivers a solution to this problem. This includes generating a breakdown of the product's quality factors as well as the users' thoughts and feelings to the product. The method of sentiment analysis can be divided into three steps. The first step is to identify the features that must be considered while evaluating views. A user who wants to buy a laptop may be interested in the battery life and durability. Hence, these might be the two factors that the reviews are evaluated for. The second step is to use the features to identify the reviews as positive or negative. This process is particularly difficult because many reviews do not express an opinion on a specific characteristic, and several reviews are false and posted on the forum by spammers and marketers. Moreover, the reality of such reviews, i.e. whether the user suggested a positive, negative, or neutral viewpoint, is unknown. We deploy binary classification algorithms for sorting. The third duty is to come up with a final breakdown for the product. On different domains, multiple types of sentiment analysis can be performed, such as fine-grained sentiment analysis, which includes helping with polarities ranging from very negative to very positive, intent-based or emotion detection, and aspect-level sentiment analysis. Similar methods using Lexicon-based approaches and machine learning-based approaches can both be used to perform sentiment analysis. Both systems have advantages and disadvantages. Aspect level sentiment analysis examines data by focusing on certain qualities or aspects.

We worked on categorization of reviews in this paper. We've used ecommerce site's data, which is comprised of structured reviews. The several challenges we have facing during review classification, as indicated above, have been addressed in two ways. First, each product review is verified before being published. Second, each review should provide a rating that may be used as the basis for comparison. The ranking is based on a five-star scale, with five stars being the best and one star being the worst. The classifiers have been applied to the generated data.

Researchers began working on semantic orientation and POS tagging in the 1950s, which led to the birth of sentiment analysis. The early 2000s were a period of rapid growth for sentiment analysis, with the highest number of articles appearing in 2004. An upgrade to the stochastic taggers utilising a new rule has been provided in work on part-of-speech tagging. The performance of the presented unique technique was improved. In addition, hidden Markov models are being used to improve sentiment analysis performance. In time series signal processes, the hidden Markov model played a critical role.

However, the first stage in every research project is to gather data, which should then be organised in a logical manner. An automated data collection method could be quite beneficial to the project. For the extraction of reviews and their associated polarities, a system called OPINE, an unsupervised approach, has been proposed. Although the retrieved data/information should be presented in a comprehensible fashion. In the work, context information in customer evaluations was uncovered. Preprocessing the data after it has been extracted is critical to the work. Aspect level sentiment analysis has given sentiment analysis a fresh perspective.

The SemEVal job, which drew the attention of many researchers, was one of the most significant contributions. The field of Aspect Level sentiment analysis has been examined in terms of its crucial and key points. Aspect level sentiment analysis is divided into two parts: classification and extraction.

Researchers presented their research on sentiment analysis algorithms for tweets. They discussed many ways to sentiment analysis, such as document level, text level, and so on. Researchers presented a study on customer Ad sharing attitudes in. They came up with the idea that sentiment analysis can help customers better comprehend their intentions for sharing ads online.

The study of emotions and sentiment analysis was the focus of researchers [13]. Natural language processing was first introduced, followed by an emotional model and other ways. They have taken publications from either DH or a computational linguistic forum into consideration for the survey.

LITERATURE SURVEY:

G.kaurand A.Singla et al., [1] used product users' reviews of products and reviews of retailers from Flipkart as a dataset, utilising a web crawler to extract comments from web pages and labelling review text by objectivity or subjectivity, as well as negative and positive buyer attitudes. Such reviews are profitable in certain ways, promising for both the consumers and the manufacturers of the products.The aim of this paper is on categorizing item reviews based on semantic meaning. They examined the principles of opinion mining, the benefits and drawbacks of previous opinion mining systems, and provided some guidance for future research. The authors propose a variety of approaches, including spelling correction in the review text, which is done to make the most sensible comment for determining the encounter of words using the Word Net dictionary, and then classifying the comment using a hybrid algorithm that combines Decision Trees and Naive Bayes algorithms. B.Pangand L.Leeet al [2]., both worked in the field of film criticism. Because big collections of online reviews are easily available, this domain is convenient to work on. Furthermore, while reviewers typically describe their overall attitude using a machine-extractable rating indicator such as the number of stars, they did not hand-label the data for supervised learning and evaluation. Their database comes from the Internet Movie Database (IMDb), which only has numeric values or ratings. The ratings are automatically extracted and classified into one of three categories: favorable, negative, or neutral. They have focused solely on determining if a sentiment is good or negative. They used the following three machine learning algorithms to do this: Nave Bayes, Maximum Entropy Classification, and Support Vector Machines (SVM).

The Explicitit Factor Model (EFM) was proposed by Y. Zhangy and G. Laiyet al.[3] to initiate suggestions while keeping a high level of prediction accuracy. They began by evoking product features (i.e. aspects) and user sentiments through phrase level opinion mining on user reviews using a new unified hybrid matrix factorization network, and then launched both recommendations and dis-recommendations based on specific item features to the user's affection and hidden features learned. The benefit of EFM over other basic algorithms on rating prediction and stop-K recommendation tasks has been established by experimental findings from several datasets. Because an explanation is also provided, the user is more likely to be influenced. Bauman, B.Liu, and A.Tuzhilin et al. [4] suggested a recommendation system that not only predicts things of affection to the user, as traditional recommendation methods do, but also a specific way of consuming the products to build the user's wisdom with those items. For example, it can encourage the user to go to a specific location (item), such as a restaurant, and order specific meals, such as Italian cuisine (an aspect of utilization). Sentiment Utility Logistic Model is the name given to their model (SULM). They put the suggested method to the test on actual sentiment reviews in three real-world apps, demonstrating that it works well in these applications by providing recommendations for the essential form that improve user knowledge.

X. Fang and J.Zhan et al. [5] provided a methodology that addressed sentiment polarity categorization, a difficult problem statement in opinion mining. They created a novel mathematical approach for emotion score computation and provided a new algorithm for negation phase identification. For sentiment polarity categorization, a feature vector generating method is proposed. Two experiments on sentiment polarity categorization were conducted on the review and sentence levels, respectively, and the implementation of three classification models was determined and compared based on the experimental results. They based their findings on Amazon's online product review data. They conducted trials on categorization at the phrase and review levels, with promising results. M.Huand B.Liuet al] presented a system for categorising reviews as unfavourable or positive based on a set of criteria.

M.Huand B.Liuet al. [6] presented a system for categorising evaluations as negative or positive based on a number of factors. The picture is created by the model in two steps: opinion direction identification and feature extraction. The method extracts all of the reviews from the inputs and stores them in the review database. The feature extraction function then pulls the "hot" features from the database that a big number of individuals have voiced an opinion about, followed by the "infrequent" ones. The retrieved features are fed into the opinion direction identification, which categorises the attribute conceptions as positive or negative. They used Association rule mining for feature mining. If an item appears more than 1% of the time in the overall number of reviews, it is deemed frequent. After then, feature trimming is done to avoid having features that are wrong. They determined the semantic orientation (i.e., positive or negative) of each opinion statement after identifying opinion features. This process involves two steps: first, they use a bootstrapping technique and the WordNet to determine the semantic orientation of each opinion word in the opinion wordlist. They then decide the idea orientation of each sentence based on those orientations. A.Ortigosa, J.M. Martn, R.M. Carro et al.,[7] proposed a classification method that takes a hybrid approach, combining machine-learning and lexical-based techniques to extract knowledge about users' sentiment polarity from user reviews, model the users' usual sentiments polarity, and detect emotional changes. The findings obtained using this method suggest that it is possible to perform sentiment analysis on Facebook with a high level of accuracy (about 83.27 percent). They have proposed new areas where sentimental analysis can play a significant role, such as e-learning, where students' opinions can serve as feedback to teachers and aid in overall improvement. Many application domains have been investigated [9-13], including stock, emoticon, and air quality prediction, as well as opinion and suggestion in the health sector, music tagging, and criminality. Y.Choi and C.Cardie et al., [14], have presented a novel learning-based technique that combines structural inference with compositional semantics in a learning procedure. Experiments conducted by them revealed that learning-based approaches that do not use compositional semantics perform worse than simple heuristics based on compositional semantics (accuracy of -89.7% vs. 89.1%). However, a strategy that combines compositional semantics and learning outperforms all other methods (90.7 percent). [15] provides a comprehensive discussion of sentiment analysis. [16-19] Various strategies for sentiment analysis have been suggested.

METHODOLOGY:

A model for better sentiment analysis has been developed using an ensemble technique to increase correctness and efficiency, and it has been tested on reviews collected for trending keywords. Overall flow of the work is shown in Fig. 1.

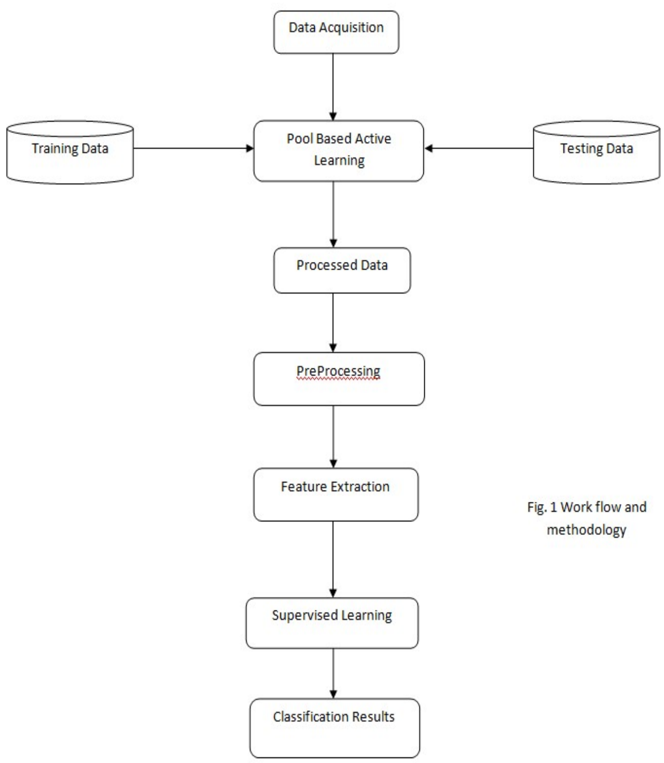

Fig. 1 Work flow and methodology

To begin, the API was used to capture the data stream and information was extracted. The extraction of aspects, which are product attributes, was then completed. The next stage is to assign ratings to the sentiments. Each phase in the work technique has its own significance. In this section, the work's methodology flow has been detailed.

We gathered data from Amazon product reviews in three categories: Electronics reviews, Cell Phone and Accessories reviews, and Musical Instruments reviews, totaling roughly 48500 product reviews. There are 21600 reviews for mobile phones, 24352 reviews for electronics, and 2548 reviews for musical instruments. We used review Text & Overall from the formats used to analyze the review polarity. An introduction of our methodology can be found here:

Research question

For this study, the following research questions have been defined:

- What features (input/output) are used in sentiment analysis?
- What methodologies are used in sentiment analysis?
- What domains have the adopted data sets addressed?
- In terms of sentiment analysis, what are the hurdles and open problems?

Search process:

We constructed a Scrapy-based web crawler to retrieve user evaluations on certain products from Amazon in order for sentiment analysis to work. Figure 2 depicts the Crawler's basic flow. An application processing interface has been created to collect data and store it. Scrapy has been used to extract data; it processes and saves data in the format. csv file format.

The data was saved in tabular format, including attributes such as the date of the review, the URL of the review, the review's rating, the user's name, and the user's review. Figure 3 depicts the total number of product reviews collected.

The term "aspect identification" refers to the process of identifying words or phrases that are related to the characteristics of the review remarks. Let's imagine the product is earbuds; the key features of an earphone are represented in Figure 4.

The human identifier is critical for identifying aspect phrases so that the aspects and their moods can be manually stored. Aspects can be identified by using a supervised approach and following aspect aggregation, which identifies terms that are synonyms of each other (for example, "battery" and "charging").

Quality assessment:

To deal with complex problems, support vector machines have been employed as a classifier, which is a supervised learning with the concept of hyper plane.

Table 1 provides some examples of bipolar words and how their polarities alter in the presence of context.

It is critical to comprehend the nature of words in order to recognize bipolar terms. When words are utilized in context, they lose their original polarity. For example, in the phrase "Dark Glossy," the word "dark" has a negative polarity, but when combined with the word "Glossy," the polarity of the entire phrase changes to Positive. Table 2 elaborates on the example.

The categorization and evaluation come next after the identification and correction of bipolar terms. As a classifier, the support vector machine was used. The RBF kernel outperformed the other two SVM kernels. The SVM classifier's operation is depicted in Figure 5.

Additional data:

- The victimization phase combinable creates a record of data that will be needed for review categorization, and a vector space model is used to do so.

- POS tagging is a type of speech tagging that allows each word of data to be tagged with a POS, such as a verb, adverb, noun, pronoun, adjective, and so on.

- Lemmatization and stemming help to reduce spatiality in words. Words like "bright," "brighter," and "brightening" are all grouped together as "bright."

- Stop word removal removes words from data that have no bearing on the data's ultimate sentiment value.

The above step is one of the most crucial steps, in which data is cleaned and stop words, among other things, are removed in order to improve the effectiveness of the results.

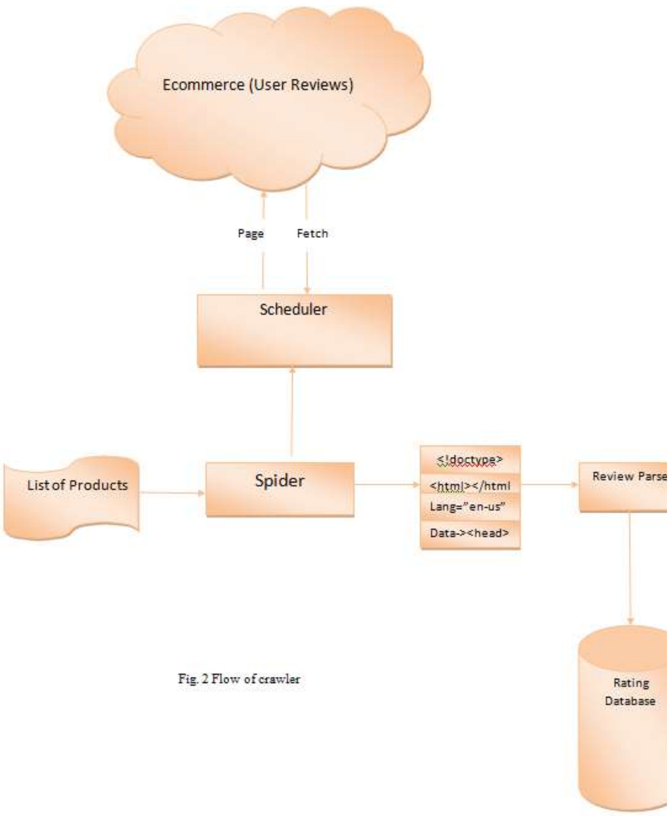

Fig. 2 Flow of crawler

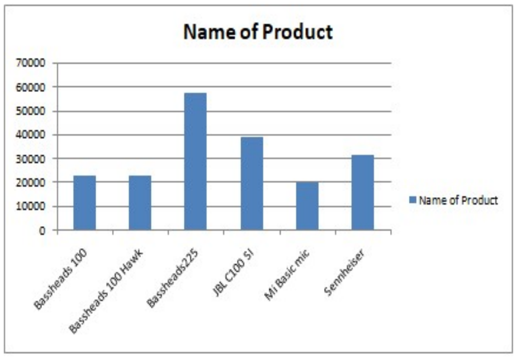

Fig. 3 showing no. of product reviews of crawled Amazon products which were collected

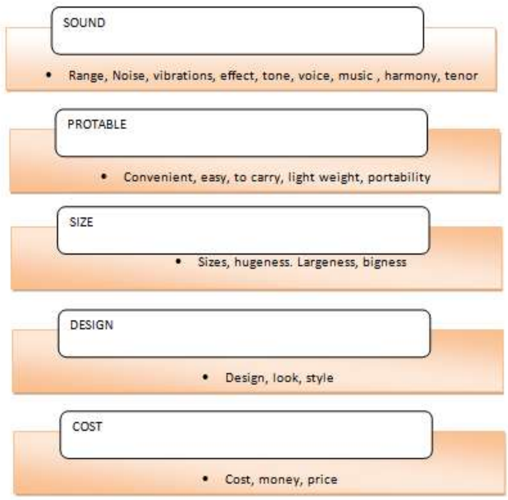

Fig. 4 Sample aspects of a product

Results and Discussion

The 34,627-review dataset was divided into three parts: a training set of size 21000 (60 percent), a validation set of size 6814 (20 percent), and a test set of size 6813. (20 percent).

Multinomial Naive Bayes, SVM with Linear Kernel, SVM with RBF Kernel, KNN-4, 5, 6, and LSTM were implemented with 4223-d input features representing review text.

KNN-5 outperforms the other two KNN models, and SVM with Linear Kernel outperforms SVM with RBF Kernel slightly. The SVM with Linear Kernel appears to have an overfitting problem, as evidenced by the significant difference between training and test accuracy. Among all of them, LSTM performs the best in terms of test accuracy.

With 50-d input features from glove dictionary, we run Gaussian Naive Bayes, SVM with Linear Kernel and KNN4, 5 6 and LSTM. KNN-5 outperforms the other 2 KNN models again. Besides, we tried data resampling on LSTM model but unfortunately it did not improve the test accuracy due to overfitting problem. It turns out that LSTM generates best predictions among all models again.

Detailed results of training and test accuracy of all models are listed in Table 1.

Models	Training Acc.	Test Acc.
Multinomial NB	75.1%	70.6%
Linear SVM	83.4%	69.6%
RBF SVM	69.7%	69.2%
KNN-4	61.7%	61.7%
KNN-5	65.5%	65.4%
KNN-6	64.9%	64.6%
LSTM	73.5%	71.5%
Gaussian NB w/ Glove	52.2%	52.4%
Linear SVM w/ Glove	68.7%	68.6%
KNN-4 w/ Glove	58.1%	57.6%
KNN-5 w/ Glove	62.6%	62.2%
KNN-6 w/ Glove	61.3%	61.6%
LSTM w/ Glove	70.1%	70.2%
LSTM w/ Glove(Resample)	85.6%	65.6%

Table 1. Performance of different models

In general, traditional input features outperform glove input features in all models. Specifically, LSTM outperforms all other models in terms of prediction accuracy. Figure 4 depicts the ranking of various models based on their test accuracy.

Discussion

During train time, we discovered that KNN required significantly more computation complexity than Naive Bayes and SVM. As with the KNN algorithm, it must calculate the distance between all of the evaluation data points and all of the training data points, which takes more time.

Furthermore, increasing the length of the dictionary had little effect on accuracy. One explanation is that as we lower the dictionary's threshold, the dictionary's length grows. But there's a problem: we only have about 40,000 reviews. When we think about it, the number of data points isn't all that much greater than the dimension of feature space. As a result, the curse of dimensionality may be at work here.

The glove mean method produces a worse result than the standard word count method.

The possible reason is that using the average weakens the individual word feature, causing the distance between different reviews to be inaccurate.

Because of the larger number of parameters, the result of LSTM is slightly better than that of other conventional machine methods. Table 1 also shows that after resampling, the training accuracy of LSTM with Glove has reached 85.6 percent. However, the test accuracy is only 65.6 percent, indicating that this model has overfitted on the resampled data due to the large number of repeated examples.

Conclusion and Future Work

In summary, we experimented with two types of features. We tried all of the algorithms mentioned in the model section for these two types of features, including Naive Bayes, SVM, KNN, and LSTM. We can see from the results that using LSTM on the first type of feature improves our accuracy on the test set. One of the primary reasons our accuracy isn't high enough is due to data imbalance. We tried resampling and different weighting techniques based on audience feedback during the poster session. But it didn't make much of a difference. Another option we haven't considered is gathering more data points from other sources. We believe this will aid us in resolving the issue of data imbalance.

REFERENCES

1. Gurmeet kaur and Abhinash Singla, "Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 1*, January 2016
2. Bo Pang and Lillian Lee, "*Thumbs up? Sentiment Classification using Machine Learning Techniques*", Proceedings of EMNLP2002.
3. Yongfeng Zhangy and GuokunLaiy, " *Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis*", SIGIR‟14 Proceedings of the 37th international ACM SIGIR conference on Research & development in informational retrieval Pages83-92
4. Konstantin Bauman, Bing Liu and Alexander Tuzhilin, "*Aspect Based Recommendations: Recommending Items with the Most Valuable Aspects Based on User Reviews*", KDD ‟17, August 13-17, 2017, Halifax, NS,Canada
5. Xing Fang and Justin Zhan, "*Sentiment Analysis using Product Review Data*", Journal of Big Data 2015
6. Minqing Hu and Bing Liu,"*Mining Opinion Features in Customer Reviews*", American Association for Artificial Intelligence2005
7. Alvaro Ortigosa, José M. Martín, Rosa M. Carro, "*Sentiment Analysis in Facebook and itsapplication to e-learning*", Department of Computer Science, Universidad Autónomade Madrid, Francisco Tomás y Valiente11, 28049 Madrid, Spain

   M Mittal, LM Goyal, J K Sethi and D J Hemanth, (2018), "Monitoring the Impact of Economic Crisis on Crime in India Using Machine Learning", Computational Economics, Springer. vol 53(4), pp. 1467-1485
8. M. Mittal, I. Kaur, S. C. Pandey, A. Verma, L M Goyal (2019) "Opinion Mining for the Tweets in Healthcare Sector using Fuzzy Association Rule", EAI Endorsed Transactions on Pervasive Health and Technology. 4.16:1-10
9. M. Shastri, S. Roy, M. Mittal, (2019) "Stock Price Prediction using Artificial Neural Model: An Application of Big Data", EAI Endorsed Transactions on Scalable Information Systems, VoL 06(20), pp. 1-8
10. Jasleen Kaur Sethi&Mamta Mittal (2019) "Ambient Air Quality Estimation using Supervised Learning Techniques", EAI Endorsed Transactions on Scalable Information Systems, VoL06(22), pp. 1-10
11. Mittal M., Arora M., Pandey T. (2020) Emoticon Prediction on Textual Data Using Stacked LSTM Model. In: Bansal J., Gupta M., Sharma H., Agarwal B. (eds) Communication and Intelligent Systems. ICCIS 2019. Lecture Notes in Networks and Systems, vol 120. Springer, Singapore
12. Kumar L., Mitra A., Mittal M., Sanghvi V., Roy S., Setua S.K. (2020) Music Tagging and Similarity Analysis for Recommendation System. In: Das A., Nayak J., Naik B., Pati S., Pelusi D. (eds) Computational Intelligence in Pattern Recognition. Advances in Intelligent Systems and Computing, vol 999. Springer, Singapore
13. Yejin Choi and Claire Cardie, "*Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis*", Department of Computer Science Cornell University Ithaca, NY 14853
14. Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., &Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, *65*, 3-14.
15. Batra, R., &Daudpota, S. M. (2018, March). Integrating Stock Twits with

sentiment analysis for better prediction of stock price movement. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-5). IEEE.

16. Sharma, M., Singh, G., & Singh, R. (2017). Stark assessment of lifestyle based human disorders using data mining based learning techniques. *IRBM*, *38*(6), 305-324.
17. Kaur, P., & Sharma, M. (2019). Diagnosis of Human Psychological Disorders using Supervised Learning and Nature-Inspired Computing Techniques: A Meta-Analysis. *Journal of medical systems*, *43*(7), 204.
18. Khan, M. A., Khan, M. A., Ahmed, F., Mittal, M., Goyal, L. M., Hemanth, D. J., &Satapathy, S. C. (2020). Gastrointestinal diseases segmentation and classification based on duo-deep architectures. *Pattern Recognition Letters*, *131*, 193-204.
19. Sharma, M., Singh, G., & Singh, R. (2019). Design of GA and Ontology based NLP Frameworks for Online Opinion Mining. *Recent Patents on Engineering*, *13*(2), 159-165.