An Integrated Approach for Amazon Product Reviews Classification Using Sentiment Analysis

Deepika Kumar
Department of Computer Science & Engineering,
Bharati Vidyapeeth's College of Engineering
New Delhi, India

Ritik Aggarwal
Department of Computer Science & Engineering,
Bharati Vidyapeeth's College of Engineering
New Delhi, India

Siddhant Raghuvanshi
Department of Computer Science & Engineering,
Bharati Vidyapeeth's College of Engineering
New Delhi, India

Shourya Chand
Department of Computer Science & Engineering,
Bharati Vidyapeeth's College of Engineering
New Delhi, India

Abstract

Sentiment Analysis plays a huge role in business analytics and situations in which text needs to be analyzed. It is used in anticipating market progression based on different news, online blogs and social media opinions. Essential part of information-gathering for market research is to find the opinion of people about the product. Many business enterprises are utilizing these opinions to perform better in the market. In this paper, the analysis is done on the Amazon product's reviews dataset. The data is organized through preprocessing and after cleaning through various techniques, some useful features are selected and sentiment analysis is done to generate a sentiment polarity. Various different learning techniques like Naïve Bayes, Linear Support Vector Machine and Logistic Regression classifiers are applied on the preprocessed data and comparison analysis is done to find the best classifier fit for the reviews data through the detailed analysis and generation of the Receiver operator characteristics curve and the comparison analysis through AUC value of different classifiers testing the precision or the accuracy. A term-doc incidence matrix is built using term frequency and inverse document frequency ponderation. The results obtained from the integrated approach of combining the sentiment polarity with the ratings to recommend a particular Amazon product proves invaluable for both the customer and seller.

Keywords - Data Preprocessing; NLP, Sentiment Analysis, Classification;

1. INTRODUCTION

Sentiment [1][3] refers to the opinion or view of a person on a specific product, topic or event. Hence Sentiment analysis, also called as opinion mining, analyze public sentiment towards some entities. Today online retailers are expanding their roots very rapidly. Not only this, they ask for user reviews on the products they sell. Users review the products on retailer's website or they can post their content on various social media platforms like social networking sites or forums etc. Hence a large pool of opinions regarding a product is readily available on the internet. If this pool is analyzed properly, it can give a lot of insight regarding the quality factors of the product. But since the information in each review is not in a defined format, it's a huge challenge to extract the relevant information from the data available. Sentiment analysis, based on the technologies of text mining and NLP, provides a way to overcome this challenge. This includes generating summary on the quality factors of the product and also the emotional orientation of the users towards that product. The process of sentiment analysis [2][4] is can be broken down into 3 crucial tasks. The first task is to figure out the characteristics for which opinions have to be analyzed. For a laptop, new users may want to know about the battery life and endurance. Hence these can be the two characteristics for which the reviews may be analyzed. The second task is to categorize the reviews as positive or negative based on the characteristics. This task is particularly challenging since many reviews do not have any

opinion about a certain characteristic and many reviews are fake and posted by spammers on the forum. Also, the ground truth of such reviews is not known i.e. whether the user implied positive, negative or neutral opinion. For categorization, we use binary classification techniques. The third task is to generate the final summary regarding the product.

In this paper we have focused on categorization of the reviews. We have used the data collected from Amazon which is nothing but structured reviews. The various challenges that we face during review categorization as mentioned above have been overcome in two ways, first, each product review receives inspections before it can be posted. Second, each review should have a rating on it that can be used as the ground truth. The rating is based on a star- scaled system, where the highest rating has 5 stars and the lowest rating has only 1 star.On this generated data we have implemented the classifiers. The various different classifiers used in this paper are the NB (naïve Bayes) [1], LR (logistic regression) [5] and SVM (support vector machine) [7].The naïve Bayes classifiers are generally used for text classification and it gives quite a good precision on small amount of dataset. For large amounts of data, naïve Bayes takes long periods of time to compute and does not give accurate results with low precision. We use the SVM [2] classifier type linear SVC (Support Vector Classifier), whose objective is to return a "best fit" hyperplane that divides, or categorizes, the data. Logistic regression [4] is used to analyze and visualize the relation between the dependent and the independent variable. The classification of reviews is done on the basis of partitioning the score out of five where a score less than three is considered a negative review while a score greater than three is considered a positive review. All The classifiers applied should perform well as only those features of the reviews data are chosen which will have an impact on the score like there view 'Summary' feature. Thus, the data size is reduced to improve the performance of the classifiers on the data.

The paper is organized as follows:

In the section II discuss about the Literature survey done in this area that gives a brief review towards related work on sentiment analysis, Section III explains about the Research Methodology used which is followed by the Results & Analysis section which is Section IV. The conclusion and future scope is discussed in Section 5.

2. LITERATURE SURVEY

G.kaurand A.Singla et al., [1] used product users review comments about product and review about retailers from Flipkart using web crawler to fetch comment from a web pages as dataset and label review text by objectivity or subjectivity and negative or positive attitude of buyer. Such reviews are profitable to some measure, promising for both the shoppers and products makers. This paper is focused on the classification of item reviews on the ground of semantic meaning. In the present study, they have analyzed the fundamentals of opinion mining, pros and cons of past opinion mining systems and supply some direction for the future analysis work. The authors have proposed completely different approaches including spelling correction in review text and the spelling correction are done to make the most sensible comment for knowing the encounter of words using Word Net dictionary and then classifying comment employing hybrid algorithm combining Decision Trees and Naive Bayes algorithm. B.Pangand L.Leeet al., [2] worked in the domain of movie reviews. This domain is convenient to work on because large collections of online reviews are readily available. Also, reviewers generally summarize their overall sentiment with a machine- extractable rating indicator like number of stars, therefore, they did not hand-label the data for applying supervised learning and evaluation. Their database source is Internet Movie Database (IMDb), where database contains only numeric value or ratings. Ratings are extracted automatically and categorized into one of these 3 categories: positive, negative, or neutral. They have concentrated only on finding the inclination of a sentiment to be positive or negative. To achieve this, they employed the following three machine learning algorithm Naïve Bayes, Maximum entropy classification and Support vector machines(SVM). Y.Zhangyand G. Laiyet al. [3] proposed the Explicitit Factor Model(EFM) to inaugurate recommendations, meanwhile maintaining a high degree of prediction accuracy. They started by evoking product features (i.e. aspects) and user

sentiments by phrase level opinion mining on user reviews using new unified hybrid matrix factorization network, then inaugurate both recommendations and dis-recommendations on the ground of specific item features to the user's affection and the hidden features learned. Experimental results of some datasets have demonstrated the advantage of EFM over other basic algorithms on rating prediction as well a stop-K recommendation tasks.Since explanation is also given, it tends to generate much greater influence on the user. Bauman, B.Liuand A.Tuzhilin et al. [4] proposed a recommendation method that not only can predict products of affection to the user as conventional recommendation method do but also specialized form of consuming of the products to further increase the user wisdom with those items. For example, it can uphold the user to visit to a specific place (item) say restaurant and also order some definitive foods there, e.g., Italian (an aspect of utilization). They have named their model as Sentiment Utility Logistic Model (SULM). This model first predicts the opinion of the user about the item based on what he/she has expressed about a specific form of the item and then determines the crucial form of the user's experience with that item. They tested the proposed method on actual sentiment reviews over three real-life practical applications and showed that their method performs strong in these practical applications by preparing recommendations about the crucial form that improve user wisdom.

X. Fangand J.Zhan et al. [5] proposed a model that tackled a very challenging problem statement of opinion mining which is sentiment polarity categorization. They proposed a new algorithm for negation phase identification and developed a new mathematical approach for sentiment score computation. A method of feature vector generation is put forward for sentiment polarity categorization. Two experiments regarding sentiment polarity categorization was respectively performed on the ground of review level and sentence level and execution of three classification models are figure out and compared on the ground of their experimental results. They have used the online products review data collected from Amazon. They conducted the experiments for sentence level and review level categorization and promising results were obtained. M.Huand B.Liuet al. [6] proposed the system for classification of reviews as negative or positive over ascertain attributes. The model performs the picture in two main steps namely opinion direction identification and feature extraction. Given the inputs, the approach first extracts all the reviews, and place them in the review database. Then the feature extraction function extracts the "hot" features from the database for which a large number of people expressed their opinion about and then it finds the infrequent ones. The opinion direction identification takes the extracted features as input and discriminates the notions of the attribute into positive or negative. For feature mining they have used Association rule mining. In this an item is considered frequent if it appears more than 1% in the total number of reviews. After that feature pruning is done to avoid incorrect features. After opinion features have been identified, they determined thesemantic orientation (i.e., positive or negative)of each opinion sentence. This process requires two steps, first for each opinion word in the opinion wordlist, they identify its semantic orientation using a bootstrapping technique and the WordNet, and then based on those orientations they decide the notion orientation of each sentence. A.Ortigosa, J.M. Martín, R.M. Carro et al.,[7] proposed a method of classification which follows a hybrid approach i.e. it combines machine-learning and lexical based techniques to extract knowledge about the users' sentiment polarity from reviews written by users and to model the users' usual sentiments polarity and detect emotional changes. The results obtained using this method show that it is practicable to do sentiment analysis in Facebook with high level of accuracy which was shown to be around 83.27%. They have proposed new areas where sentimental analysis can play a huge role like in e-learning where the opinions of students can serve as a feedback for the teacher and help in overall improvement. Many application areas like stock, emoticon and air quality prediction, as well as opinion and recommendation in health sector, music tagging and crime sector have explored [9-13].

Y.Choiand C.Cardie et al., [14] have presented a novel learning based approach that integrates structural inference with compositional semantic into a learning procedure. Experiments conducted by them have shown that learning based methods without the use of compositional semantics perform worse as compared to simple heuristic based on compositional semantics (accuracy of 89.7% v/s 89.1%). A method that combines compositional semantics

with learning however performs the better than all other methods (90.7%).A detail review of sentiment analysis has been discussed in [15].Different techniques for sentiment analysis has been discussed [16-20]

3. RESEARCH METHODOLOGY

Data Collection

The study is done on the product reviews data obtained from AWS which has a large repository of varietal public datasets. From September to November2016, 5.6 million reviews were collected in which various products belong to different categories: food, beauty, book, electronics, home, appliances. Those online reviews were posted by o v e r 4.4 million reviewers (customers) towards 20,062 products. Each particular review of a product consists of various attributes like the product ID, reviewer ID, score, time of the review, helpfulness and the review text. The ratings are based on a five-star scale, from one to five with no halves or quarters. These ratings are all compiled in a separate column 'score' in the reviews data.

Data Loading and Preprocessing

The data cleaning is done so as to use only the useful data and remove the redundant or the unnecessary data. The cleaning is done through the following steps:

1) Stemming: It is a form of data pre-processing with natural language processing. It is done to reduce the different words to its root form.
2) Stop words removal: This step removes the redundant words so as to reduce the size of the data. The stop words filter out the words in the stop word list like if, are, how and hence prepare the data for further analysis.
3) Tokenization: This is a basic step where the sentence is broken into words or tokens. Textual data comprises block or string of characters called tokens. The review comments are separated as tokens and used for further processing.
4) Textual transformation and Classification: The score is transformed from the rating of each review to the sentiment categorization as positive or negative.
5) Evaluation: The data after preprocessing is evaluated by applying the linear SVC, Logistic regression and Bernoulli naïve bayes classifiers and the accuracy and recall value is analyzed and further comparison analysis done.

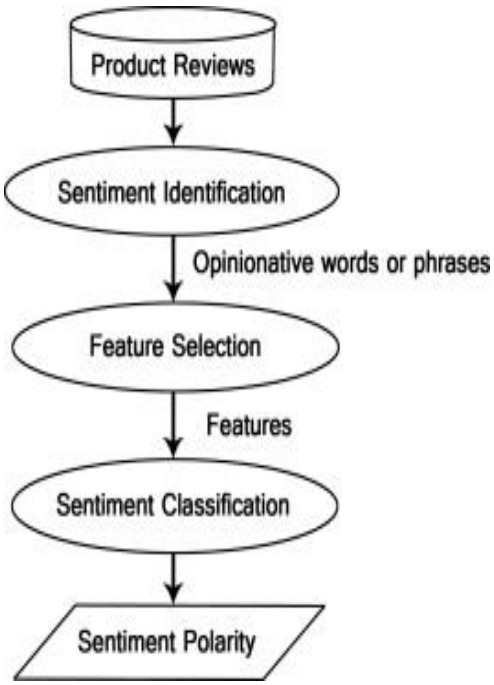

Fig. 1. Steps used in sentiment classification

Sentiment Analysis and Feature Selection : The sentiment analysis is done on the review text by matching the positive and negative review text words with the comment text and generating a short summary text. The features selected from the preprocessed data are short summary text and the ratings score which form the basis for building the integrated model for classifying the product into multi-variate classes of recommendation.

Sentiment Classification and polarity

Classification is a process of determining the form of data points .The SQLite dataset is used which will fetch the score and the comment text. The data set will be split into training set and a test with a test ratio of 0.2. The data is trained for 80% data and then tested on remaining 20% data. Different classifiers are applied to predict the recommendation's sentiment and comparison analysis through ROC Curve and analyzing the AUC value of different classifiers done to predict which classifier is the best "fit" for the data set.

Parameters for evaluation: For classification, TP, TN, FP, FN are used in the analysis and generation of ROC curve. True positive means, which are truly classified as the positive terms. True Negative means, which are truly classified as the Negative terms. Other evaluation measures like precision, recall, specificity and accuracy can easily be calculated from these four variables.

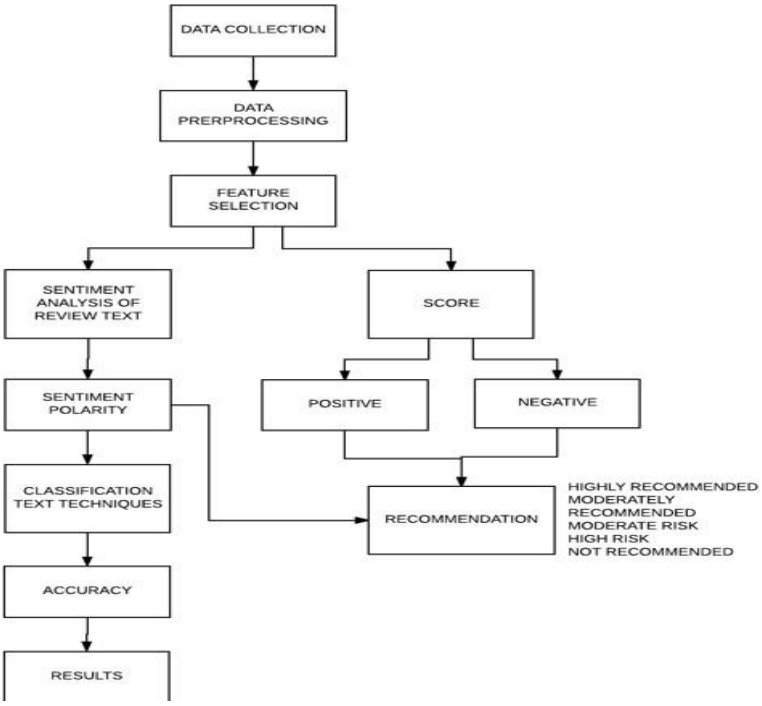

Fig.2.IntegratedSystemArchitectureforProduct Recommendation

Accuracy – A definitive measure of the classification performance of an algorithm on the transformed data is the accuracy. Accuracy is the proportion of correctly classified examples to the total number of examples, while error rate uses incorrectly classified instead of correctly. Accuracy though should be carefully used when using skewed data. This is because when one class occurs significantly more than the other, the classifier might get higher accuracy by just labeling all e examples as the dominant class then what it gets when it tries to classify some with the other class.

Precision and recall – For text mining and other text analysis field like information retrieval, precision and recall are two most widely used metrics for evaluating performance. They can be seen as extended versions of accuracy, and by using a combination of these measures the problem with skewed data for classifiers dissipates.

Initially, the score attribute data is classified as positive and negative through partitioning the score. Above 3 classifies the review as positive and below 3 as negative. The sentiment polarity generated from the review text is combined with the partitioned score to form the integrated structure for classifying the product into different classes of recommendation: highly recommended, moderately recommended, moderate risk, high risk and not recommended.

4. COMPARISON ANALYSIS AND RESULTS

The sentiment analysis on a product is performed by matching the positive and negative words text file with the review text of each product. The compilation generates a short summary text which when merged with the score gives the recommendation of the particular product which is useful for both customers and sellers. Various learning method classifiers like SVC, logistic regression, NB are applied to evaluate which model is the best fit for the predicting the sentiment.On the analysis of the performance of the three classifiers, SVM, logistic regression and Bernoulli naïve bayes, we analyze that the area under the curve value of both SVM and logistic regression is better than the naïve bayes classifier. As a reduced size data set is used, the Bernoulli naïve bayes performs better than expected giving an accuracy of 0.75. For large data though this value will go down. The SVM and the logistic regression give the AUC value of 0.77 and 0.78 respectively. Thus, the logistic regression outperforms the other two marginally. Hence it is best "fit" for thedata.

Fi

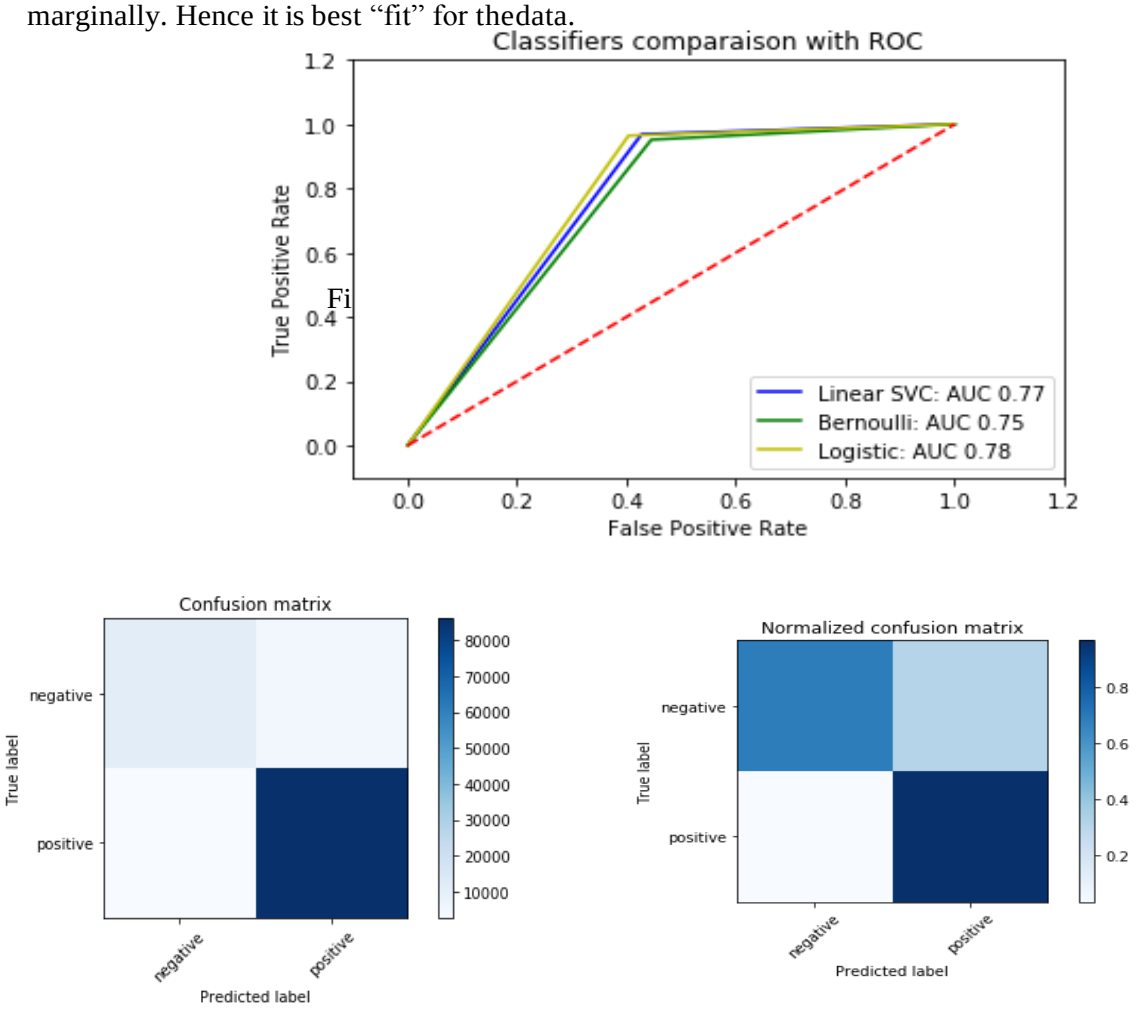

Fig. 4. Confusion matrix

Table 1: Comparison Analysis

Score	Positive Sentiment	Negative Sentiment
1	Not recommended	Not recommended
2	High Risk	High Risk
3	Moderate Risk	Moderate Risk
4	Recommended	Moderate Risk
5	Highly recommended	Moderate Risk

Conclusion and Future Scope

The sentiment analysis on a product is performed by matching the positive and negative words text file with the review text of each product. The compilation generates a summary text which when merged with the score gives the recommendation of the particular product which is useful for both customers and sellers. Various learning method classifiers like SVC, logistic regression and NB are applied to evaluate which model is the best fit for the predicting the recommendation sentiment. The recommended product is useful for online shoppers as well as product manufacturers and sellers. It also enables the manufacturers and sellers to work on reducing the fragilities with the negative sentiment product, underlying the various problems and methods to resolve the design and intricacies with the product.

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
8. 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

9. 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
10. 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
11. Jasleen Kaur Sethi&Mamta Mittal (2019) "Ambient Air Quality Estimation using Supervised Learning Techniques", EAI Endorsed Transactions on Scalable Information Systems, VoL 06(22), pp. 1-10
12. 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
13. 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
14. 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
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.