Survey On Feature Selection for Data Mining and its Application in Opinion Mining

Sumanth S[*] & Siddarama S[†]

ABSTRACT

Sentiment Analysis (SA) and opinion mining is used for the systems of business intelligence in analyzing public opinion towards various brands and implementing market strategies. Machine learning aims at developing the algorithm in such a way that the performance of a system is optimized with past data or experience. In the case of SA, the feature selection method is used for the identification of different goals such as reduction of cost of computation, avoidance of over-fitting, and enhancing the accuracy of classification of the model. The methods of feature selection can reduce their original feature sets by the removal of irrelevant features for the classification of text sentiment and their accuracy. The paper surveys various feature selection techniques available in the literature. The study shows that feature selection significantly improves the classification of the sentiments, but it depends on the technique adopted and the number of features selected.

[*] Department of Computer Science, Smt. V.H.D. Central Institute of Home Science(A), #463, Seshadri Road, Bengaluru – 560 001, Karnataka, India; sumanth81s@gmail.com

[†] Department of Computer Science, Sri Bhagawan Mahaveer Jain College, Robertsonpet K.G.F.– 563 122, Karnataka, India; siddu_hulimani@yahoo.co.in

1 INTRODUCTION

The increase in the use of online activities and the Internet (such as clickstreams, micro-blogging, blogging, social media communications, e-commerce, online transactions, ticketing, surveillance, conferencing, and chatting) has resulted in the need for data. This has to be extracted, transformed, loaded, and analyzed in large amounts of both unstructured and structured data and is popularly known as big data. This data is analyzed by means of combining data mining, text mining, and web mining-based techniques. Large amounts of information that is related to customer reviews and opinions can be very cumbersome while analyzing the approaches in order to procure an opinion summary that is generalized. There are several websites, forums, blogs, and social network platforms that express opinions that are used to understand the opinion of the public. This includes an opinion on reputation monitoring, product preference, marketing, company strategies, social events, and political movements [1].

In order to get this accomplished, academicians and research communities have been rigorously working on SA for about fifteen years. SA refers to a computational study consisting of opinions, attitudes, emotions, and sentiments that are expressed using texts toward a particular entity. SA is also known as attitude analysis, appraisal extraction, review mining, or opinion polling. It is the task that detects, extracts and classifies sentiments and opinions that are expressed using textual input. It is useful in achieving goals such as observation of public mood with regard to political movements, prediction of movie sales, customer satisfaction measurement, and market intelligence. Reviews, evaluations, and sentiments are evident owing to the development of e-commerce [2].

Data pre-processing techniques [3] have a crucial role in cleaning up the irrelevant punctuations, characters, or words that are not feasible for the machines' interpretation. Hence, Natural Language Processing (NLP) will define diverse techniques that can be picked

as per the use case. In certain cases, the order in which these techniques are employed is also quite significant. In general, the data or text pre-processing techniques will handle the conversion of the raw data into a comprehensible structure in which the importance is given primarily to the keywords present in the text that highlight the paragraph or sentence's context. 1. Sentence Segmentation, 2. Change to Lower Case, 3. Tokenisation, 4. Parts-of-Speech Tagging, 5. Stopwords Removal, 6. Removal of Punctuations, 7. Stemming, and 8. Lemmatization isone of the frequently applied techniques of pre-processing. Feature extraction refers to representing features in the form of vectors so as to make them more comprehensible to the machine. Every feature extracted by these techniques extract will have a vector form representation prior to it being given as input to the classifier models. 1. Named Entity Recognition (NER), 2. Bag-of-words Model (BoW), and 3. Term Frequency-Inverse Document Frequency (TF-IDF) are a few techniques of feature extraction.

Feature selection-based methods have been classified to be attributed to the evaluation algorithms or subset evaluation algorithms. The former has all features ranked in an individual manner with a weight assigned to each of the features in accordance with every degree of relevance that it had to the target feature. The latter, on the other hand-selected feature subsets and ranked them on the basis of certain other evaluation criteria. The methods of attribute evaluation did not measure any correlation among features and, therefore, could yield subsets that had redundant features. These techniques of feature selection were grouped into three: the filter, the wrapper, and the embedded methods. Each algorithm of feature selection used one of the three techniques of feature selection. This is also known as the subset selection in machine learning, where the subset of a feature will be available from the chosen data [4].

The methods of feature selection have been comprehensively divided into three: filter techniques, wrapper techniques, and embedded techniques. Each of these elements in selection makes use of any of the three strategies. The framework of machine learning was an integrated system with programs from current data that can predict new observations. Machine learning further

deals with the study of systems from data as opposed to following the instructions that were programmed. The technique was further used in a varied range of tasks. The opinion normally originates from a certain state of mind in our daily lives. This expression can be a negative comment or an appraisal. Certain techniques can identify and further predict sentiments from the text, such as Lexicon, Machine Learning techniques, and NLP. Machine learning can design and also develop a computational algorithm. It was a term coined by computer scientist Arthur Samuel in the year 1959 when working on a pattern recognition algorithm in the IBM Lab. Machine learning originated from artificial intelligence as a method of pattern recognition in which NLP has a major role to play as an artificial intelligence branch [5].

Machine learning generally provides facilities to computers without any explicit programming. Machine learning also has statistics in the form of optimization, which makes it very feasible. It is categorised into Unsupervised, Supervised, and reinforcement types. For the unsupervised regression, it is data-driven clustering that is taken into consideration with reinforcement algorithms that learn to react to different environments. For supervised regression and classification, the approach is in detail and covers all data completely. There are three major domains of machine learning that are detailed thoroughly to support different applications and techniques. The use of this method in the building of applications comes with different banners, such as data mining, spam filtering, character recognition, and search engines [6].

Different classifiers used for opinion mining are recommended here, and among them, the rule-based classifiers, Support Vector Machines (SVM), decision trees, and the Naïve Bayes are widely used. The work done in SA for classification is as below [7]:

1. Word or Phrase Sentiment Classification: Today, the classification of Word Sentiment is the very basis for document and phrase classification.

2. Document Sentiment Classification: The approach of supervised machine learning is widely used for predicting the document and its overall sentiment. The main objective of the task was the classification of every review document

to express either a negative or a positive sentiment on an object.

3. Sentence Level Sentiment Classification: As the researchers found it extremely coarse to calculate document-level sentiment, they further investigated certain approaches for determining the actual focus of every sentence. Many other researchers also studied sentence-level sentiment classification, which denotes the classification of every sentence as a form of expressing either a positive or negative opinion.

This work reviews feature selection and meta-heuristic methods for opinion mining. The rest of the paper is organized into the following sections. Section two discusses related works available in the literature, and section three presents the conclusion.

2 RELATED WORKS

Riaz et al., [8] conducted a SA based on customer reviews using real-world data at the phrase level in order to identify the preferences of customers by means of making an analysis of various subjective expressions. The strength of sentiment words was then computed for every single expression, and clustering was applied to place the words in different clusters on the basis of their intensity. The results of this technique were further compared to the star ranking provided in the dataset. It was observed that there was a major change in the results. A visual representation of the results was provided in order to obtain a clear customer preference insight and their behaviour that helps decision-makers make better decisions.

Kang et al., [9] proposed another new method of SA that was based on the Hidden Markov Models (TextHMMs). This used a word sequence for training as opposed to the predefined sentiment lexicon. The text patterns that represent sentiment by means of an ensemble TextHMMs were used, and the method defined the hidden variables of the TextHMMs using semantic cluster information. It took the co-occurrence of words into consideration to compute sentiment orientation using fitted TextHMMs. For reflecting the diverse patterns, a new ensemble of the TextHMM-

based classifiers was used, and the method proved to be superior to the existing ones in classifying implicit opinions. It also was practical in real-life datasets of online reviews.

Kaladevi and Thyagarajah[10] made a proposal for finding opinions in social media using Integrated Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) Deep Neural Network(ICNN-LSTM-DNN). The proposed technique was adaptable to the mechanism and could investigate the posts on social media and also the facts obtained from the online Twitter messages. The ICNN-LSTM-DNN-based SA had been applied to the tweets connected to the Indian Election in 2019. This approach outperformed other techniques in terms of F-Measure, recall, precision, and accuracy.

Basiri and Kabiri[11] made an exhaustive investigation considering machine learning with methods based on the lexicon. Another new method was proposed for the problem of rating prediction in the Persian language. The effect of this machine learning component, feature selection, combination level, and normalization were investigated. The results of the experiment for larger datasets with a total of 16,000 Persian customers' reviews demonstrated the ability to attain a better level of performance compared to the Naïve Bayes, which was a lexicon-based method. Also, the experiments demonstrated that it could be used in polarity detection as well.

Eo and Lee [12] made a new proposal of a novel model of opinion mining that was based on a combination of the methods of Feature Selection (FS) along with Word Embedding to Vector (Word2vec) and the Bag-of-words (BOW). These FS methods were adopted for the work and were the (Correlation-based FS) and the Information Gain (IG). For the purpose of choosing an FS method that was optimal, there were several other classifiers that are ranging from the Logistic Regression (LR), Naive Bayesian Network (NBN) Neural Network (NN), to the Random Forest (RF), the Random Subspace (RS), and Stacking ST. Empirical results combined with electronics along with kitchen datasets proved that the LR and the ST classifiers combined along with IG duly applied to BOW features gave the best opinion mining performance. The results

with the laptop and the Word2vec features gave the best opinion mining performance.

Hemalatha and Selvi[13] extracted biomedical opinions from Twitter, consisting of several features that are required for the classification of opinions. But these datasets also contain certain weak and irrelevant features that can influence prediction accuracy. Without another algorithm for feature selection, it may be challenging for the classification techniques to identify the patterns of features. The actual purpose of feature selection was not just identifying the subset from its original feature sets but also bringing down overheads of computation in data mining. In this feature selection approach, the Shuffled Frog Leaping Algorithm (SFLA) was used to optimize the feature subsets to increase predictive accuracy. This is used as a feature selector and also to generate the KNN classifier, SVM, Naïve Bayes, and other feature subsets. The results of the experiment proved that the Naïve Bayes showed better levels of accuracy when the features were chosen from the SFLA.

Jain et al., [14] further proposed Particle Swarm Optimization (PSO), which was hybridized using the Neutrosophic Set for the generation of a ternary classifier. Sentiment for a large text was grouped in the form of a ternary value which can be positive, negative, and sometimes neutral. The method was also well-suited for the classification of texts of a large size. Such hybridization was not dealt with in literature. The positive, negative, or neutral texts were produced from the review comments of the movie plots or the products. The method was extremely useful in classifying documents such as comments of review on works of research from various researchers, news to find sentiment polarity, political analysis, and so on.

Jiang et al., [15] proposed another methodology to be used for dynamic modeling of the preferences of customers that were based on customer reviews made online that include Dynamic Evolving Neural-Fuzzy Inference Systems (DENFIS) and Opinion mining. This was adapted for analyzing all online reviews and further performing SA on these reviews for different periods of time. The DENFIS method keeping these mined time series data with the settings of product attributes for the reviewed products, performed

dynamic modeling for customer preferences. This was for the preferences of customers in connection with their variance of errors or meant relative errors. Additionally, the DENFIS approach was also able to provide fuzzy and crisp outputs that were not realized by means of using the other conventional DENFIS and ANFIS approaches.

Sangam and Shinde [16] developed another new framework used for opinion mining. This included a method of feature selection known as the Most Persistent Feature Selection (MPFS) along with a Genetic Algorithm (GA) based method of optimization that was used for feature set optimization. The MPFS further makes use of features and the information gained for the review of documents. This feature set has produced the GA technique to get an effective feature set for its sentiment classification. The SVM algorithm was employed to classify sentiments of various reviews that were expressed in the text using the method of optimization and feature selection. These classifier models further generated an acceptable level of performance for accuracy compared to the other models.

Ernawati and Yulia[17] made use of the Naïve Bayes along with the GA for classifying the text to review the online fashion companies. This resulted in classifying the text as either positive or negative. The measurements were based on the NB and its accuracy even before the addition of the Genetic Algorithms (GA) and also after its addition. A validation by using the 10-fold cross-validation was made with a ROC curve and confusion matrix for measuring accuracy. The work aimed at increasing the NB algorithm's accuracy, and the results proved that it was evident that the GA was able to improve overall accuracy.

Kurniawati and Pardede [18] made a proposal for PSO to choose the attributes that were appropriate from the documents and make use of the SVM as its classifier. After this, a SA system was employed for electing the Governor of West Java. The results of the experiments proved that the accuracy of the proposed system was about 94.80%, and its Area-Under-Curve (AUC) value was 0.98. Results also proved that there were major improvements when the PSO with Information Gain was used.

Jain et al., [19] made another proposal of a novel hybrid framework called the "Senti-NSetPSO" for making an analysis of texts of a larger size. The Senti-NSetPSO includes two different classifiers, which are the binary and the ternary-based PSO with the Neutrosophic Set. The method was well-suited for classifying a large number of texts with a size of above 25 kb. The Swarm size that was generated from a larger text provides a measurement to implement the convergence of the PSO. This approach was further tested in the case of large-sized texts that were obtained from the Subjective Dataset, Polarity, ac1IMDb, and the Blitzer. The method also established a good correlation between the Neutrosophic Set and SA. On Polarity Dataset, Blitzer, and aclIMDb, there was a satisfactory level of accuracy. The accuracy of the ternary classifier had improved significantly.

Akyol and Alatas [20] presented opinion mining and SA that were used for analyzing the problems of a social network. Also, in the initial stages, SA was considered to be a problem of SA, which was a multi-objective problem. The Social Impact Theory-based Optimization Algorithm and the Whale Optimization Algorithm are intelligent optimization algorithms that are adapted for problems of SA. Also, a memory feature will be integrated into a Social Impact Theory-based Optimization Algorithm to obtain results that are effective. These results were compared to thirty-three supervised learning algorithms in real IMDB, Amazon datasets, and Polarity. For evaluating the performance of these results, the MCC, F-Measure, precision, recall, and accuracy percentage are used. On examining the results, it was observed that these metaheuristic optimization algorithms were more successful in SA.

Osmani et al., [21] proposed an Imperialist Competitive Algorithm (ICA) that was hybridized to improve for the Artificial Bee Colony (ABC) and its exploitation. The methods further improved the results of the evaluation for both discrete and continuous baseline methods. The next method was able to improve the methods of optimization. Feature selection was one problem of optimization as the choice of suitable feature subsets can be crucial to improve the efficacy of the classifiers. Thus, in order to focus on the methods of feature selection, the key issue is quite important. For the purpose

of this work, there were various discrete versions of the methods proposed that were introduced and could be used for feature selection problems of scoring that were successfully evaluated. For this work, another problem known as the cold start was introduced, and this presented a new solution that had a great impact on the proposed method and its efficiency in managing the problems of feature scoring. About 16 UCI datasets and 2 Amazon datasets were used to evaluate this problem.

Alarifi et al., [22] introduced another novel big data and machine learning technique to assess the process of SA. All noise in the data was removed by making use of pre-processing in data mining. All effective features were obtained from the pre-processed data with a greedy method to choose the optimal features that were processed using an optimal classifier known as the Cat Swarm Optimization-based LSTMNN (CSO-LSTMNN). These classifiers analyze all sentiment-related features based on the behaviour of the cat when examining their features. The results that were obtained using the greedy feature with the CSO-LSTMNN algorithm and PSO algorithm were duly compared; the CSO-LSTMNN was able to outperform the PSO with regard to decreasing the error rate and increasing the rate of accuracy.

Padmavathy and Mohideen[23] primarily focused on the prediction of drug satisfaction levels among patients with a two-pass classifier. This Two-pass classifier refers to a combination of the SVM with the ANN (SVMNN). Initially, customer reviews were collected from the domains of healthcare. All the important features of these reviews were collated. The two-pass classifier is applied to this to predict whether the review is negative or positive. The proposed approach and its performance are analyzed using the F-measures, precision, and recall. The results of the experimentation proved the proposed system to have better results compared to the other methods available.

Ahamed et al., [24] deployed fuzzy-based machine learning for accomplishing a fine-level SA for large online opinions, which was done by assimilating fuzzy linguistic hedges that influence opinion descriptors. There are different algorithms of Machine Learning, such as the Ada Boost, Naïve Bayes, SVMs, and the K-Nearest Neighbour, which are used for classification purposes. J Soup had been implemented to gather different web opinions that were

subject to the initial task of processing. This was later applied to tagging and stemming. Such fuzzy-based methodologies were then investigated for the laptop datasets and Mobiles and were compared to other state-of-the-art methods. This was done to demonstrate an upper indication of an accuracy of around 94.37% using Kappa indicators with lower rates of errors. All outcomes of the investigation concluded that the proposed method was used efficaciously in Sentimental analysis in helping online decisions.

Kumar et al., [25] made another proposal for efficient SA methods, which was effectively implemented using three different processes: (a) the creation of ontologies to extract semantic features, (b) Word2vec to convert the processes corpus, (c) a CNN used for opinion mining. To tune the CNN parameters, PSO was utilized. The experimental results proved that the technique was able to outperform the other techniques that were state-of-the-art and also yielded 88.52%, 85.63%, 94.30%, and 86.03% in terms of accuracy, recall, precision, and F-measure.

Keyvanpour et al., [26] made a proposal for a machine learning technique and lexicon known as the OMLML for use in various social networks. The advantage of the proposed method was that it could address all challenges at the same time and could determine the polarity of the target words with a lexicon-based method along with textual features of sentences and words. The machine learning algorithms classified the mapped feature space. The quantitative and qualitative results of the experiment proved that the mapping of data within a new space could decrease the cost of training. The proposed method's performance was quite acceptable with regard to the accuracy, runtime, and F-measure.

Asghar et al., [27] proposed another integrated framework consisting of a set of new heuristic patterns used for aspect extraction. It made use of a hybrid sentiment module for classification using additional support from negations, summary generators, and intensifiers. The evaluation of the performance of this aspect-based opinion mining system was able to outperform other methods with regard to precision, recall, and F-measure. The results were compared, and observed that the technique was more efficient in summary generation for online reviews of products, classification, and extraction of aspect sentiment.

Li et al., [28] further proposed another approach to opinion mining that was based on the concept of Kansei Engineering (KE) along with machine learning for extracting the measures of users and their affective responses to the products. The Ridge Regression (RR), SVM, Support Vector Regression (SVR), Classification and Regression Tree (CART), and Multi-Layer Perceptron (MLP) algorithms were applied. Another experiment was conducted to illustrate this method. The results proved that the SVM+SVR performed the best. The recall, precision, and F1 score was more than 80% while classifying soft-hard attributes using a mean square error that was the smallest. On the basis of the method proposed, the manufacturers and designers were aware of the responses of customers with regard to products and review texts for designing products.

Khan et al., [29] further investigated the applicability of various algorithms of supervised machine learning in connection to classifying comparative reviews. Here, different types of algorithms were applied to perform multi-class classification for comparative reviews of various classes. The results proved that the Random Forest was able to outperform other classifiers.

Ghosh et al., [30] further proposed another accurate model for SA to review restaurants, movies, and products from Yelp, IMDP, and Amazon. Using classifiers like the decision tree, SVM, and logistic regression, the authors were able to classify the reviews as either positive or negative with better accuracy.

Bhalla and Bagga[31] proposed the RB-Bayes approach that was based on the Bayes theorem that removes the zero-likelihood problem. The approach was compared to the other existing approaches, such as the SVM and the Naïve Bayes. The technique was demonstrated with better outcomes and helped in better dataset analysis. When the proposed approach was tried on some datasets, the accuracy was higher, and the RB Bayes had a precision of about 83,333.

Tiwari et al., [32] proposed another machine learning approach for classifying the tweets of passengers with regard to their airplane services that help in understanding the emotions and their patterns. A Random Forest (RF) with Logistic Regression was adopted to classify every tweet as to whether it is positive, negative, or neutral. An evaluation of the computed real data helped in demonstrating

both methods and their accuracy, which was 80%. Table 1 shows the summary of the literature reviews.

S. No	Authors	Techniques	Merits	Demerits
1	Jiang et al., [15]	DENFIS and ANFIS	Better performance	Dynamic modeling
2	Sangam and Shinde [16]	MPFS and GA	Performance for accuracy	Feature set optimization
3	Ernawati and Yulia [17]	NB and GA	To improve overall accuracy	Reviews
4	Jain et al., [19]	Senti-NSetPSO	Satisfactory level of accuracy	A large number of texts
5	Akyol and Alatas [20]	Social Impact Theory-based Optimization Algorithm and the Whale Optimization Algorithm	Meta-heuristic optimization algorithms were more successful in SA	Multi-objective problem
6	Osmani et al., [21]	ABC-ICA	Efficiency	Feature selection problems
7	Alarifi et al., [22]	CSO-LSTMNN	Improving the efficiency of the system	Large dataset volumes
8	Padmavathy and Mohideen [23]	SVMNN	Better results	Effects of drugs
9	Kumar et al., [25]	CNN and PSO	Efficient SA methods	Non-dominant Pareto front
10	Li et al., [28]	KE and machine learning algorithms	Better performance	Online reviews

Table 1 Summary of Literature Reviews

3 CONCLUSION

This paper has explored the impact of feature selection methods on the performance of sentiment analysis in information systems. Sentiment analysis has a wide range of applications, including categorizing reviews, summarizing reviews, extracting synonyms and antonyms, and tracking opinions in online discussions. A comparative study was conducted on feature selection methods for sentiment categorization of online movie reviews. The main contribution of this research is the evaluation of the accuracy of various feature selection and machine learning methods in sentiment analysis. NLP, a subset of Artificial Intelligence, will handle how the machines will comprehend as well as interpret human language. A computer will understand everything in the form of numbers instead of words. Hence, it will be better to examine which techniques of pre-processing and feature extraction will have to be deployed on a human language such that upon its number-format conversion, it will be feasible for interpretation by the computer. Future works can explore the performance of techniques of feature selection on diverse machine learning classifiers and also assess the model for cross-domain sentiment analysis.

REFERENCES

1. Kumar, B. S., & Reddy, B. (2016). An Analysis on Opinion Mining: Techniques and Tools. Indian Journal of Research, 5(8).

2. Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-Based Systems, 89, 14-46.

3. Tabassum, A., & Patil, R. R. (2020). A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing. International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 06, 4864-4867.

4. Samsudin, N., Puteh, M., Hamdan, A. R., &Nazri, M. Z. A. (2013). Immune based feature selection for opinion mining. In Proceedings of the World Congress on Engineering (Vol. 3, pp. 3-5).

5. Varma, R. R., Nagesh, Y. N., Umesh, I. M., & Pradeep, N. C. Opinion Mining using Machine Learning. International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS) Volume VI, Issue XII, December 2017 | ISSN 2278-2540

6. Gamal, D., Alfonse, M., M El-Horbaty, E. S., & M Salem, A. B. (2019). Analysis of Machine Learning Algorithms for Opinion Mining in Different Domains. Machine Learning and Knowledge Extraction, 1(1), 224-234.

7. Patel, V., Prabhu, G., &Bhowmick, K. (2015). A survey of opinion mining and sentiment analysis. International Journal of Computer Applications, 131(1), 24-27.

8. Riaz, S., Fatima, M., Kamran, M., & Nisar, M. W. (2019). Opinion mining on large scale data using sentiment analysis and k-means clustering. Cluster Computing, 22(3), 7149-7164.

9. Kang, M., Ahn, J., & Lee, K. (2018). Opinion mining using ensemble text hidden Markov models for text classification. Expert Systems with Applications, 94, 218-227.

10. Kaladevi, P., &Thyagarajah, K. (2019). Integrated CNN-and LSTM-DNN-based sentiment analysis over big social data for opinion mining. Behaviour & Information Technology, 1-9.

11. Basiri, M. E., &Kabiri, A. (2020). HOMPer: A new hybrid system for opinion mining in the Persian language. Journal of Information Science, 46(1), 101-117.

12. Eo, K. S., & Lee, K. C. (2019). Investigating Opinion Mining Performance by Combining Feature Selection Methods with

Word Embedding and BOW (Bag-of-Words). Journal of Digital Convergence, 17(2), 163-170.

13. Hemalatha, S. M., &Selvi, C. K. (2018). Feature Selection for Opinion Mining Using Shuffled Frog Leaping Algorithm. International Journal of Engineering and Computer Science, 7(02), 23656-23662.

14. Jain, A., Nandi, B. P., Gupta, C., &KumarTayal, D. (2019, March). A hybrid framework based on PSO and neutrosophic set for document level sentiment analysis. In International Conference on Information Technology and Applied Mathematics (pp. 372-379). Springer, Cham.

15. Jiang, H., Kwong, C. K., Kremer, G. O., & Park, W. Y. (2019). Dynamic modelling of customer preferences for product design using DENFIS and opinion mining. Advanced Engineering Informatics, 42, 100969.

16. Sangam, S., & Shinde, S. (2018, May). A Novel Feature Selection Method Based on Genetic Algorithm for Opinion Mining of Social Media Reviews. In International Conference on Information, Communication and Computing Technology (pp. 167-175). Springer, Singapore.

17. Ernawati, S., &Yulia, E. R. (2018, August). Implementation of The Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies. In 2018 6th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-5). IEEE.

18. Kurniawati, I., &Pardede, H. F. (2018, October). Hybrid method of information gain and particle swarm optimization for selection of features of SVM-based sentiment analysis. In 2018 International Conference on Information Technology Systems and Innovation (ICITSI) (pp. 1-5). IEEE.

19. Jain, A., Nandi, B. P., Gupta, C., & Tayal, D. K. (2020). Senti-NSetPSO: large-sized document-level sentiment analysis using Neutrosophic Set and particle swarm optimization. Soft Computing, 24(1), 3-15.

20. Akyol, S., &Alatas, B. (2020). Sentiment classification within online social media using whale optimization algorithm and social impact theory-based optimization. Physica A: Statistical Mechanics and its Applications, 540, 123094.

21. Osmani, A., Mohasefi, J. B., &Gharehchopogh, F. S. (2020). Sentiment Classification Using Two Effective Optimization Methods Derived from The Artificial Bee Colony Optimization and Imperialist Competitive Algorithm. The Computer Journal.

22. Alarifi, A., Tolba, A., Al-Makhadmeh, Z., & Said, W. (2020). A big data approach to sentiment analysis using greedy feature selection with cat swarm optimization-based long short-term memory neural networks. The Journal of Supercomputing, 76(6), 4414-4429.

23. Padmavathy, P., &Mohideen, S. P. (2020). An efficient two-pass classifier system for patient opinion mining to analyze drugs satisfaction. Biomedical Signal Processing and Control, 57, 101755.

24. Ahamed, S., Danti, A., & Raghavendra, S. P. (2019, March). Feature Based Fuzzy Framework for Sentimental Analysis of Web Data. In 2019 International Conference on Data Science and Communication (IconDSC) (pp. 1-5). IEEE.

25. Kumar, R., Pannu, H. S., &Malhi, A. K. (2020). Aspect-based sentiment analysis using deep networks and stochastic optimization. Neural Computing and Applications, 32(8), 3221-3235.

26. Keyvanpour, M., Zandian, Z. K., &Heidarypanah, M. (2020). OMLML: a helpful opinion mining method based on

lexicon and machine learning in social networks. Social Network Analysis and Mining, 10(1), 1-17.

27. Asghar, M. Z., Khan, A., Zahra, S. R., Ahmad, S., &Kundi, F. M. (2019). Aspect-based opinion mining framework using heuristic patterns. Cluster Computing, 22(3), 7181-7199.

28. Li, Z., Tian, Z. G., Wang, J. W., & Wang, W. M. (2019). Extraction of affective responses from customer reviews: an opinion mining and machine learning approach. International Journal of Computer Integrated Manufacturing, 1-16.

29. Khan, A., Younis, U., Kundi, A. S., Asghar, M. Z., Ullah, I., Aslam, N., & Ahmed, I. (2019, April). Sentiment classification of user reviews using supervised learning techniques with comparative opinion mining perspective. In Science and Information Conference (pp. 23-29). Springer, Cham.

30. Ghosh, S., Hazra, A., & Raj, A. (2020). A Comparative Study of Different Classification Techniques for Sentiment Analysis. International Journal of Synthetic Emotions (IJSE), 11(1), 49-57.

31. Bhalla, R., &Bagga, A. (2019). Opinion mining framework using proposed rb-bayes model for text classification. International Journal of Electrical & Computer Engineering (2088-8708), 9(1).

32. Tiwari, P., Pandey, H. M., Khamparia, A., & Kumar, S. (2019). Twitter-based opinion mining for flight service utilizing machine learning. Informatica, 43(3).