

Emotion Classification through Product Consumer Reviews

Shahid Salim¹, Zeshan Iqbal, Javed Iqbal

Department of Computer Science, University of Engineering and Technology, Taxila, 47080, Pakistan

¹Corresponding author: E-mail: Shahid Salim (shahid.salim@students.uettaxila.edu.pk)

Abstract- Currently, the world is moving towards digitalization. Ecommerce is enamoring strength in this digitalized world through the accessibility of products reachable for customers. The recent work in the field of emotion classification based on consumer feedback reviews is gaining popularity worldwide. However, the research lacks to provide better accuracy in terms of pre-processing and feature extraction. Since, to determine consumer emotions in terms of their choices, feature extraction and pre-processing plays an important role to enhance the accuracy rate to determine their choices for product marketability and to identify consumer behavior in selecting product features resulting in increased product demands. The review of related literature shows that the products accuracy rate can be improved through refining the dataset and by executing proper pre-processing techniques. Examining its feature extraction by applying different algorithms, which required a lot of pre-processing of reviewed data to make the features more valuable. This research provides an analysis of cars, hotels and mobile datasets and studies sentiment classification with different machine learning Approaches. First, to minimize the noise of the dataset was undergoing with pre-processing steps including punctuation, stop words, null entries and duplication removal. In the next step, extract the feature by two different methods: Count Vectorizer and bag of words. After that bidirectional encoder representation from transformers (BERT) algorithm were applied on three datasets to predict the results. By applying BERT Classifier on cars dataset 98.5% accuracy were founded, by applying same algorithm on hotels dataset 98.3% accuracy were founded, and for Mobile dataset 98.7% accuracy were founded.

Index Terms-- Bag of Words, BERT, Sentiment Analysis, Text Classification, Online Product Reviews

I. INTRODUCTION

Currently, the world is becoming digitalized. Ecommerce is enamoring strength in this digitalized world through the accessibility of products reachable for customers. Moreover, the ecommerce website permits the people to convey what they think and feel. Actually, people are progressively relying on the experiences of other customers. Our opinion and buying direction are impacted by the experience of others and their opinion about products. We generally ask others about their opinion to get the benefit from their experience; hence, the importance of reviews has mature. However, it is almost impossible for customers to read all such reviews; therefore, sentiment analysis represents a vital role in analyzing them.

This research proposes a sentiment analysis to predict the polarity of Amazon mobile phone dataset reviews using supervised machine learning algorithms. Sentiment analysis helps a customer to make their buying decision based on the experience of others. Further, it will help companies to improve the quality of their products by knowing customers' opinions and needs [1]. We can classify the reviews, comments and document in two types. These are numeric sentiment and categorical sentiment. Numeric sentiments contain common example of numeric sentiment is rating system in ecommerce sites. Using this rating system company judge the response of peoples. Categorical sentiment is the technique to classify the comment or review in different categories. These categories are binary (positive and negative), ternary (positive, negative and neutral) and multiple categories (Sad, Anger, happy etc.) [2].

Sentiment Analysis is a major task of NLP (Natural Language Processing). Opinion has developed into a critical component of enhancing the efficacy of all human activities and products. There are two kinds of review websites: generic review websites and unique review websites. Generic websites include Amazon.com, epinions.com, and rottentomatoes.com, as well as tripadvisor.com, yelp.com, and other domain standards. Both of these review sites play a significant role in our decision-making process. These options include purchasing a camera, a cell phone, and other products, as well as making investments in any product. By incorporating the results of this input into the product development process, the overall quality of the product can be improved. Human decision-making is affected by sources other than the Internet, such as friends, relatives, and others. Figures 1.1 and 1.2 illustrate the positive and negative reviews for the Apple iPhone 7 on Amazon [3].



FIGURE 1. POSITIVE REVIEW



FIGURE 2. NEGATIVE REVIEW

II. RELATED WORK

Sentiment Analysis is a more advanced type of content mining (SA). SA is a term that refers to the automatic collection of assumptions, predictions and subjectivity in the paper. During the review of late figures several improvements and alternative SA implementation will be addressed and quickly implemented. The authors in [4] focused on online product reviews sentiment analysis. They collected food dataset from different websites. The data values were divided into Contents-based, Grades-based, along with Collaborations based (CLB) setting as of the dataset. They used deep learning modified neural network to get positive, negative and neutral reviews based on the deep learning modified neural network. The authors in [5] researched on Amazon mobile Reviews to predict people thoughts. They collected dataset from amazon website. For data preprocessing they used (Stop Words remove, Duplication remove, Null Values Remove, Punctuation Remove). For feature extraction they used two methods bag of words and TF-IDF. They used Bert algorithm for Text Classification to predict results.

Text mining, according to Shazia H et al [6] is used to analyze people's thoughts. Emotional analysis is the way people's thoughts are collected. This article classifies the emotional repercussions of film reviews. For feature extraction they used TF-IDF. They used Support Vector algorithm for Text Classification to predict results. Khanvilkar et al. [7] researched conclusions that differ from normal emotional speech. The aim of the calculation is to decide the extreme written emotion of a person. Concept assessments are helpful in developing product ideas. According to industry audits, drugs may be administered to another customer. Evaluation surveys on real-world object pages are performed to determine the prevalence and estimates. The majority of participants were split into two groups: positive and negative. Any structure is simplified by restricted study of ordinal groupings. The computer uses SVM and Random Forest AI sparingly. The mechanism proposed will be used to test the customer survey segment which uses common problems. Based on their own personal experience, customers make decisions. Partha Mukherjee et al, [8] researched on different Product Reviews to predict people thoughts. They collected dataset from different website. For data preprocessing they used (Stop Words remove, Duplication remove, Null Values Remove, Punctuation Remove). For feature extraction they used two methods Word2vec. They used Recurrent Neural Network algorithm for Text Classification to predict results.

R Rathore and others edited it Opinion mining is increasingly popular in the field of study, due to the difficulties of dealing with analytical problems and the potential benefits in a variety of real-

world applications. In particular, mining is an important topic in this area for future-oriented evaluation. We propose a new approach for extracting opinions from theory-based item audits, using sound assessment data and trees to provide clear and verifiable perspectives in this paper [9]. Text classification is important in many applications for text mining. Khanvilkar and colleagues examined the increasing attractiveness of people towards social media networks and web pages for e-commerce. As social media is growing, all Fortune 500 companies are now using sentiment analyses. Feeling analysis has been a popular topic in natural language research for a long time. This article discusses the performance and logistical regression of machine learning algorithms such as the Multinomial Nave Bay, the Clan SVM, the Decision Tree and the Random Forest. The sentiment analysis results of the above classifiers are analyzed using comparative tablets. In certain cases, random forests can produce significant results. The most versatile and straightforward learning algorithm is supervised machine learning. Random Forest has an outstanding performance in the proposed scheme. User goods guidelines are generated using the polarity determined by different algorithms [10]. Shreya Desai et al. [11] compiled the largest social media resource. Opinions on social media sites have an enormous influence. The content analysis can be used to evaluate the value of a number of objects. They collected dataset from different social website. For feature extraction they used TF-IDF. They used Multi Perception Layer Classifier for Text Classification to predict results.

Hu et al. [12] propose the hybrid approach for automated customer service analysis involving data mining and text mining with an emphasis on hotel reviews. The key four stages of the method are selection, extraction and aggregation. Useful experience and evaluations are often taken into account during the product creation process. The author prefers supervisory preparation with examples of training for the description of the product and its explicit characteristics. The assessments and aggregation combine and summaries the features and terminology of the commodity. The Research Aspect establishes dependencies between product characteristics by defining key consumer satisfaction determinants and applying association rules to the entire product by means of decision treaties.

III. METHODOLOGY OF RESEARCH

Make sure that the latest solution exceeds existing alternatives to achieve a more effective approach in every field of sentiment analysis transitions. In a process called sentimental analysis, the methodology proposed was compared with established approaches with a higher level of response and credibility, especially in terms of extraction aspects. The detail of the various phases is shown in the following subsections.

A. DATASETS DESCRIPTION

Data collection is a very important part that played a vital role in any research. Data collection is based on the different types of research. One of the types is, Quantitative research which needs Quantitative data like numerical values needs calculations, and the other one is Qualitative research which needs survey-used data, observational data etc. According to our research (Reviews

classification), we required text type data in supervised form like in the form of CSV or excel_ which also need Some pre-processing steps and calculations for data analysis by using machine learning. The dataset choose for this research is a collection of review dataset of different hotels and cars. A complete text review includes the fields extracted and the favorites, dates and names of the authors. The automotive data collection is detailed in Table I.

TABLE I
DATASETS DESCRIPTION

Dataset Name	Number of Reviews
Car Reviews Dataset	14200
Hotel Reviews Dataset	8500
Mobile Reviews Datasets	3700

The proposed methodology is given in Fig. 3. It consists of several steps, such as preprocessing, extraction of aspect and opinion, refinement of aspects and product improvement outcomes. It is used with a relation classifier to propose product improvement. The detailed description of all the steps and processes involved in proposed methodology are described as under

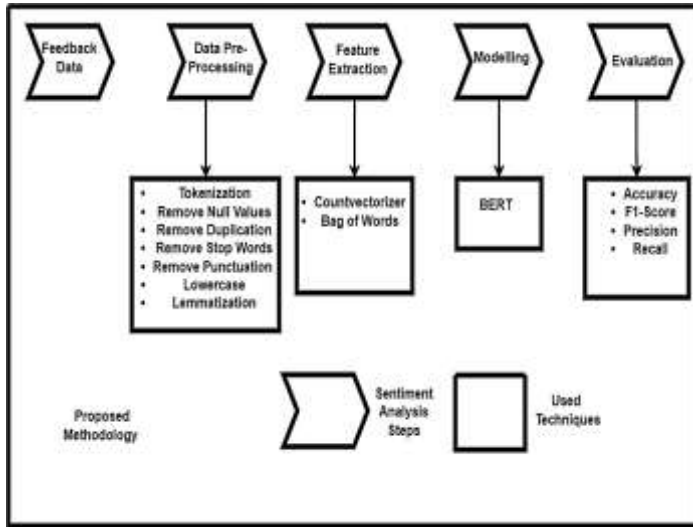


FIGURE 3. METHODOLOGY OF RESEARCH

B. DATA PRE-PROCESSING

The degree of preprocessing is shown in Algorithm. All this algorithm is simple from input to output and reduces, standardizes, sticks to a redundant substance and transliterates to extremes of a positive yet insufficient presence. Figure 4 shows preprocessing systems at all stages, including tokenization, reducing, standardization, stemming, erasure of irrelevant data and transliterations. The figure shows the entire procedure, which begins with feedback or input into the database and is still pre-processed in accordance with figure 4 If each process is

completed, the data is cleared and all model functions will be performed at the final stage.

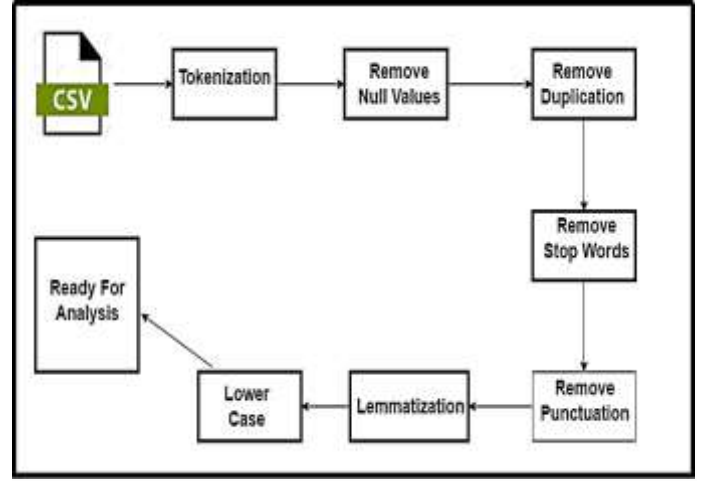


FIGURE 4. DATA PRE-PROCESSING

i. TOKENIZATION

Tokenization is the process of assigning a meaning to something. Tokenization is the process of converting text into tokens and vectors. Unnecessary tokens can also be filtered out more quickly. Paragraphs or sentences, for example, are included in the text. To tokenize the feedback in this case, we use terminology. The following step in the pre-processing process is a smaller case procedure, which transforms upper case to lower case to clean up the data and remove noise. The tokenization procedure is the next step [13].

ii. LOWECASE

It is a common preprocessing phase that reduces the entire document to one of the best text preprocessing types. It makes a major contribution to the predicted results coherence for most NLP and text mining problems [14].

iii. REMOVE STOPWORDS

In this phase that we remove stop words from the entire document. It makes a major contribution to the predicted results coherence for most NLP and text mining problems.

iv. REMOVAL OF IRRELEVANT CONTENT

Separate markings, special features and numerical tokens are removed so that the impression is not added that only alphabetical characters remain. This procedure must be performed as tokenized deletion terms are divided. The next step is the translation in the target language of a word from the source language and phonetics. The approach to transliteration is applied [15].

v. REMOVE PUNCTUATIONS

It is a common preprocessing phase that we remove punctuation from the entire document. It makes a major contribution to the predicted results coherence for most NLP and text mining problems.

C. FEATURE EXTRACTION

The goal of our proposed system is to extract consumer feedback data such as opinions on products and present this information in efficient and effective way. Consumers express their opinions in review sentences with single words or phrases. We need to extract

these opinion words or phrases in efficient way. We use two methods to extract features from datasets which are given below.

i. COUNT VECTORIZER

Count Vectorizer is also called term frequency, which determines the frequency-based value of the token. It saves the occurrence of every separate token in the text, the higher the value of the token with the higher the occurrence of the token in the text. The approach of Count Vectorizer also uses the technique of bag of words (Bow) to assess the corresponding text, every text is based on multiple bags of words that represented the occurrence of the specific word in specific sentence [18].

ii. BAG OF WORDS

Bag of Words is a model, that represented the text in the form of a vector of its words, ignoring the order and structure of the sentence but store frequency of all of the words included in that sentence or entries of the dataset, which used for the training of classifier. In (Bow), when the entries of the dataset are processed through the Bow model, Bag of Words (Bow) uses count Vectorizer as its submodule to present the matrix of the group of entries in the dataset which is available in python as SK-learn library. It allocates the specific number of occurrences of a word in that specific entry. This process is repeated for all entries of a specific dataset and stores the frequencies of all the words separately. Example of Bag of Words is given below [19].

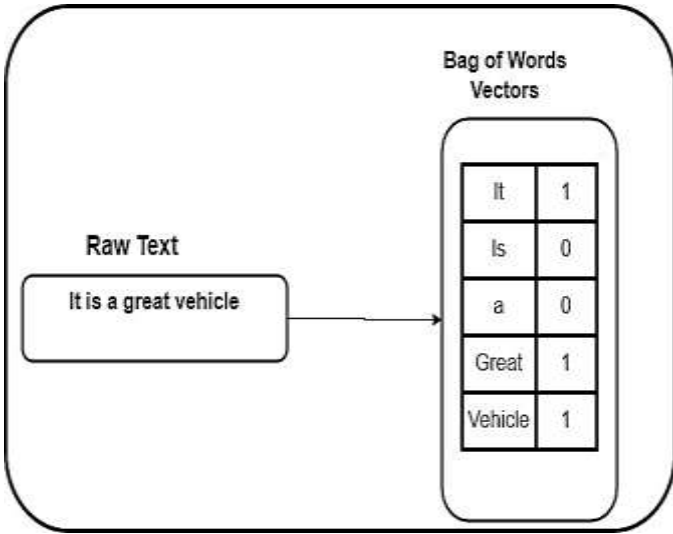


FIGURE 5. BAG OF WORDS EXAMPLE

D. MODELING

It is necessary to identify ways to improve the procedure after the final results are obtained. Relational classification has been used for the evaluation of class probabilities in a subset of entities with class labels. The method of classification is based on two fundamental assumptions. The first class names are considered to be individual classes within a single structure. Secondly, it shows people who belong to the same social class in one or more dimensions. The performance of the classifier will be weak if every entity and mark is unknown. Bidirectional Encoder Representation from Transformers is a multilayered encoder. It is used for text classification for large datasets. In our proposed

system we use Bidirectional Encoder Representation from Transformers for the better results.

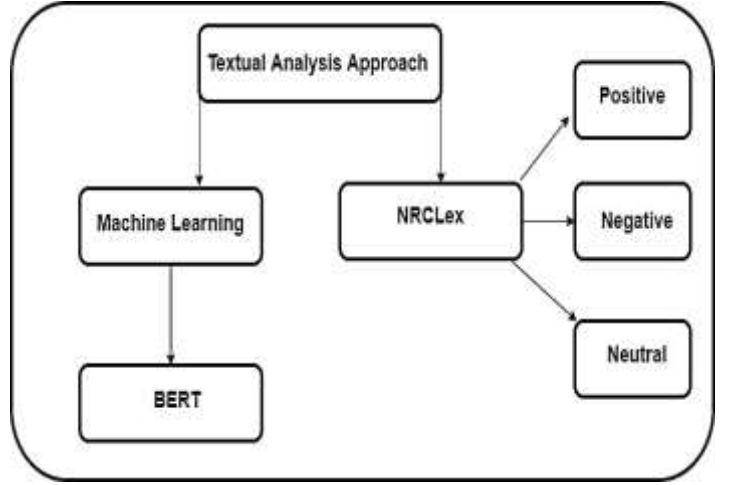


FIGURE 6. MODELING

E. PERFORMANCE EVALUATION PARAMETERS

We used Performance metrics to measure the performance of learning algorithm, there are different performance measures, some common (Accuracy, Recall, F1 Score, and Precision) that we used to analyses the performance of our f algorithms regarding our work. The measures that we discuss are:

1. Accuracy
2. Recall
3. F1-Score
4. Precision

TABLE II
EVALUATION METRICS

Evaluation Metrics		Actual	
		Positive	Negative
Prediction	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Accuracy measure is used to check that how many predictions a classifier makes are predicted correctly. However, this measure is not much convinced regarding the dataset, which is not balanced due to the issues of biasedness of predictions towards the category that is high in frequency and predict the other categories wrongly.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

where TP is True Positive, TN is True Negative, FN is False Negative, and TP is True Positive. These terms of evaluation metrics are used to evaluate the accuracy of prediction over the actual class of that entry. Similarly, Recall, Precision and F1-Score are used to measure the performance of the classifiers.

These are much better than that of accuracy to determine the performance of the algorithms when the dataset is not well balanced, and their formulas are given as,

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ Score = 2 * \left(\frac{(Precision * Recall)}{(Precision + Recall)} \right) \quad (4)$$

We discuss all these measures by confusion metric of all best performing algorithms and measure the performance [21-29].

IV. RESULTS AND DISCUSSIONS

Following the achievement of the desired result, recommendations for process improvement are required. At this point, the relational classification has been used to generate class probabilities that are based solely on items of the same type with known class labels. If things are separated or if no labels are known, it is possible that the classification will not be conducted correctly. This project shows the execution of a variety of machine learning model Bert, with different feature extraction approaches for a text classification task. Further, we used the pre-trained BERT model and fine-tuned it for the sentiment analysis Cars, Hotels, Mobile reviews dataset. The proposed methodology yielded the following results:

A. EXPRIEMNTS ON DATASETS

The procedures that were offered yielded encouraging results. The dataset for Cars, Hotels and Mobile was analyzed in the research. Table given below present the findings of a suggested methodology. When evaluating the efficacy of the methodology, Bert classifiers were used to determine its effectiveness. We achieved the high accuracy when we use Bert classifiers. For the Car, Hotel and Mobile dataset, the proposed approach is utilized in conjunction. Performance of the suggested methodology is summarized in Table II in terms of accuracy.

TABLE III
EVALUATION METRICS

Dataset	Accuracy	F1-Score	Precision	Recall
Car Dataset	0.9833	0.9824	0.9832	0.9833
Hotel Dataset	0.9842	0.9816	0.9840	0.9842
Mobile Dataset	0.9872	0.9854	0.9870	0.9872

B. SENTMENT ANALYSIS

Opinion mining is also named as Sentiment analysis. It is the method of Natural Language Processing to analyze the text either its showing positivity, negativity, or natural. It is an expression, emotion, opinion, or attitude towards a specific topic then these expressions are extracted and classified against that specific text.

There are different methods to extract the emotion from the body of the text. We used NRCLex techniques to analyze our data of Cars Reviews, Hotel Reviews and Mobile Reviews to evaluate the Review percentage of either it's positive or negative. We applied this technique after the pre-processing step of our methodology to get more efficient results. Now, here we discussed the results based on positive and negative Car Reviews. 79% of Reviews are positive or having positive emotions in their words and the remaining 21% are negative reviews. From Hotel Review 79% of reviews are positive or having positive emotions in their words and the remaining 21% are negative reviews. . From Mobile Review 69% of Reviews are positive or having positive emotions in their words and the remaining 31% are negative reviews. Figure 7 shows the Sentiment Analysis of our datasets.

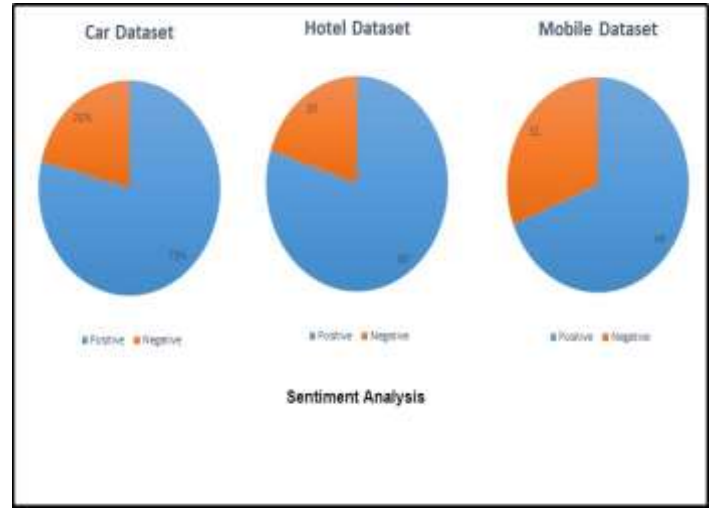


FIGURE 7. SENTIMENT ANALYSIS

V. CONCLUSION

A large number of experiments have been carried out in order to evaluate the effectiveness of the proposed methodology. Each of the strategies outlined above is used to improve the efficiency of the extraction process. The output of each measure is the culmination of a variety of evaluating perspectives. Comparing the proposed method to the multiple present strategies, the proposed method produces excellent results. As demonstrated, it is effective across all sectors. When compared to the previous technique, the proposed methodology shows extremely promising results.

VI. FUTURE WORK

For future work, we are planning to use word2vec for feature extraction with our models and to detect fake reviews. In the future, we can enhance our work by adding more for large datasets to get more accurate results with high accuracy. Preprocessing can also play its role to improve the accuracy when we use stemming, tokenization, which can also help to get the more effective features by applying more different feature extraction techniques and played a vital role in the improvement of our proposed methodology. Our proposed methodology can help to improve the quality of the product.

ACKNOWLEDGMENT

All Praises are for Almighty Allah, who enabled me to complete this work successfully and my greatest respect to Muhammad (P.B.U.H) his last Prophet. I trust it is a presumptuous benefit to explain my wonderful appreciation and a significant feeling of commitment to my reverend Supervisor whose relentless direction, consolation and backing have consistently been a wellspring of inspiration.

REFERENCES

- [1] Aljuhani, Sara Ashour, and Norah Saleh Alghamdi. "A comparison of sentiment analysis methods on Amazon reviews of Mobile Phones." *Int. J. Adv. Comput. Sci. Appl* 10, no. 6, pp.608-617, 2019.
- [2] Tayal DK, Yadav SK, "Fast retrieval approach of sentimental analysis with implementation of bloom filter on Hadoop," In 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT) 2016 Mar 11 pp. 14-18.
- [3] Nawaz, Asif, Ashfaq Ahmed Awan, Tariq Ali, and Muhammad Rizwan Rashid Rana. "Product's behaviour recommendations using free text: an aspect based sentiment analysis approach." *Cluster Computing* 23, no. 2, pp. 1267-1279, 2020.
- [4] Sasikala, P., and L. Mary Immaculate Sheela. "Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS." *Journal of Big Data* 7, no. 1, pp. 1-20, 2020.
- [5] Arwa S. M. AlQahtani, "Product Sentiment Analysis for Amazon Reviews" international Journal of Computer Science & Information Technology (IJCSIT), vol 13, no 3, June 2021
- [6] Shaziya, Humera, G. Kavitha, and Raniah Zaheer. "Text categorization of movie reviews for sentiment analysis." *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 4, no. 11, pp. 11255-11262, 2015.
- [7] Khanvilkar, G., & Vora, D., "Product Recommendation using Sentiment Analysis of Reviews: A Random Forest Approach, *Int. J. Eng. Adv. Technol.* 2019.
- [8] Rajeswari, B., S. Madhavan, Ramakrishnan Venkatesakumar, and S. Riasudeen. "Sentiment analysis of consumer reviews—a comparison of organic and regular food products usage." *Rajagiri Management Journal*, 2020.
- [9] Mate, C. Product aspect ranking using sentiment analysis: A survey. *International Research Journal of Engineering and Technology*, vol. 3, no. 01, pp. 126–127, 2015.
- [10] Khanvilkar, G., & Vora, D.. Sentiment Analysis for Product Recommendation Using Random Forest. *International Journal of Engineering & Technology*, vol.7, no.3.3, pp. 87–89, 2018.
- [11] Desai, S., & Han, M. Social Media Content Analytics beyond the Text: A Case Study of University Branding in Instagram. *Proceedings of the 2019 ACM Southeast Conference*, 2019, pp. 94–101.
- [12] Hu, Y.-H., Chen, Y. L., & Chou, H.-L., "Opinion mining from online hotel reviews, a text summarization approach," *Information Processing & Management*, vol. 53, no. 2, pp. 436–449, 2017.
- [13] Saito, Y. and Klyuev, V., "Classifying user reviews at sentence and review levels utilizing Naïve Bayes," In *2019 21st International Conference on Advanced Communication Technology (ICACT)*, 2019, pp. 681-685.
- [14] Wassan, Sobia, Xi Chen, Tian Shen, Muhammad Waqar, and N. Z. Jhanjhi. "Amazon product sentiment analysis using machine learning techniques." *Revista Argentina de Clínica Psicológica* vol. 30, no. 1, pp. 695, 2021.
- [15] Bahrawi, Nfn. "Sentiment Analysis Using Random Forest Algorithm-Online Social Media Based." *Journal of Information Technology and Its Utilization* 2, no. 2, pp. 29-33, 2019.
- [16] Fikri, Mohammad, and Riyanarto Sarno. "A comparative study of sentiment analysis using SVM and SentiWordNet." *Indones. J. Electr. Eng. Comput. Sci.* 13, no. 3, pp. 902-909, 2019.
- [17] N. Tamara and Milievi, Comparing sentiment analysis and document representation methods of Amazon reviews, 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), pp. 000283--000286, 2018.
- [18] Ogada, Kennedy, Waweru Mwangi, and Wilson Cheruiyot. "N-gram based text categorization method for improved data mining." *Journal of Information Engineering and Applications*, vol. 5, pp. 35--43, 2015.
- [19] Guellil, Imene, and Kamel Boukhalfa. "Social big data mining: A survey focused on opinion mining and sentiments analysis." In *2015 12th international symposium on programming and systems (ISPS)*, 2015, pp. 1–10
- [20] Yang, Qijnuan, Yanghui Rao, Haoran Xie, Jiahai Wang, Fu Lee Wang, and Wai Hong Chan. "Segment-level joint topic-sentiment model for online review analysis." *IEEE Intelligent Systems* 34, no. 1, pp. 43-50. 2019.
- [21] Rathore, Rahul. "Rule based approach for aspect extraction from product reviews." PhD diss., Dhirubhai Ambani Institute of Information and Communication Technology, 2018.
- [22] Ireland, R., & Liu, A., "Application of data analytics for product design: Sentiment analysis of online product reviews. *CIRP Journal of Manufacturing Science and Technology*, 23, pp. 128–144, 2018.
- [23] Li, S., & Li, Y., "A Sentiment Analysis of Online Reviews Based on the Word Alignment Model: A Product Improvement Perspective. In 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2018, pp. 2226–2231.
- [24] Nawaz, A., Asghar, S., & Naqvi, S. H. A., "A segregational approach for determining aspect sentiments in social media analysis," *The Journal of Supercomputing*, vol. 75, no. 5, pp. 2584–2602, 2019.
- [25] Santosh, D. Teja, K. Sudheer Babu, S. D. V. Prasad, and Abhishek Vivekananda. "Opinion mining of online product reviews from traditional LDA Topic Clusters using Feature Ontology Tree and Sentiwordnet." *IJEME* vol. 6, pp. 1-11, 2016.
- [26] Yang, Qijnuan, Yanghui Rao, Haoran Xie, Jiahai Wang, Fu Lee Wang, and Wai Hong Chan. "Segment-level joint topic-sentiment model for online review analysis." *IEEE Intelligent Systems* 34, no. 1, pp. 43-50, 2019.
- [27] Tago, Kiichi, Kosuke Takagi, Seiji Kasuya, and Qun Jin. "Analyzing influence of emotional tweets on user relationships using naive bayes and dependency parsing." *World Wide Web* 22, no. 3, pp. 1263-1278, 2019.
- [28] Aggarwal, Charu C. "Opinion mining and sentiment analysis." In *Machine learning for text*, pp. 413-434. Springer, Cham, 2018.
- [29] Jagdale, Rajkumar S., Vishal S. Shirsat, and Sachin N. Deshmukh. "Sentiment analysis on product reviews using machine learning techniques." In *Cognitive Informatics and Soft Computing*, pp. 639-647. Springer, Singapore, 2019.