Sentiment Analysis on e-commerce product reviews using Machine Learning and Natural Language Processing

Prof. Pradeep N. Fale[1]
[1]Department of Information Technology,
Priyadarshini College of Engineering, Nagpur, India.
[1] pradeep.fale@gmail.com

Nikita Swain[2], Pranay Rahangdale[3], Pranay Waghmare[4], Harshwardhan Bagde[5], Nikhil Dhakate[6]
[2,3,4,5,6]UG Student, Department of Information Technology, Priyadarshini College of Engineering, Nagpur, India.
[3]rahangdale.pranay010@gmail.com

Abstract— **Reputation-predicated trust models are widely utilized in e-commerce applications, and feedback ratings are aggregated to compute sellers' reputation trust scores. The "all good reputation" quandary however is prevalent in current reputation systems – reputation scores are ecumenically high for sellers and it is difficult for potential buyers to cull trustworthy sellers. Predicated on the optical discernment that buyers often express opinions openly in free text feedback comments, we have proposed CommTrust, a multi-dimensional trust evaluation model, for computing comprehensive trust problems for sellers in e-commerce applications. Different from subsisting multidimensional trust models, we compute dimension trust scores and dimension weights automatically via extracting dimension ratings from feedback comments using Natural Language Processing (NLP).**

Keywords- *NLP Natural Language Processing, sentiments, product reviews, e-commerce, machine learning.*

e-commerce are working on analyzing product reviews to enhance their business on an immensely colossal scale, and those who are not working on it because of trepidation of getting negative replications are losing a sizably voluminous deal every day. So, to avail them to analyze the sentiments of customers we are working on sentiment analysis of e-commerce product reviews utilizing Natural Language Processing (NLP).
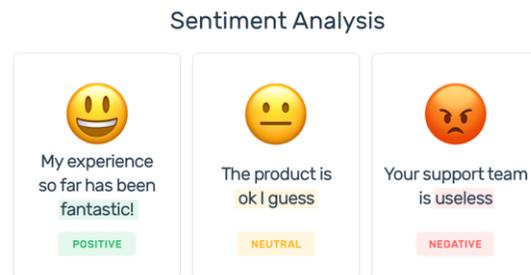

Figure 1 Product review analysis

I. INTRODUCTION

Precise product review becomes more paramount when it comes to evaluating customer gratification about the product to calculate the estimated selling of the product as this generation is mostly on the internet and trust online selling portals to buy their circadian needs. People look for product reviews for buying a diminutive thing in the cyber world and search for the product which has more positive ratings and reviews. Even most of the e-commerce platforms are probing for more reviews from the customers to avail incipient customers buy that product for them as they believe in sharing reviews so that the product could be sold in an astronomically immense quantity. No matter how top brand one is running or just commenced but what matters to the buyers is what kind of reviews the product got. Product reviews do play a crucial role in your e-commerce business as far as credibility is concerned. Credibility is one of the key elements that decide the prosperity of your brand in the long run. All the top brands of

II. LITERATURE SURVEY

After Ali et al. [1], the feedback comments are considered as a source where buyers express their opinions more voraciously and openly. The analysis of feedback comments on eBay and Amazon reveals that even if a buyer gives a positive rating for a transaction, s/he still leaves comments of mixed opinions regarding different aspects of transactions in feedback comments. Unlike subsisting trust models (including the one utilized on eBay) where explicit transaction feedback ratings (positive or negative) are habituated to compute overall trust scores for sellers.

Suraj Gund et al., [2], developed the proposed system to guide the users in a felicitous way so that they can cull the correct seller between the different available sellers on E-commerce. There are different sellers as per their accommodations and quality they provide the product to users. Users can give

reviews for particular dimensions of sellers such as quality, accommodation, distribution, etc. Review can be negative or positive and to calculate the final score of each seller in particular dimensions we utilized a different technique. Authors represent the opinion mining, aspect miming for the review mining, or feedback mining. Xiuzhen

Zhang et al. [3], predicated on the visual examination that buyers often express opinions openly in free text feedback comments, we propose CommTrust for trust evaluation by mining feedback comments. Authors have proposed a multidimensional trust model for computing reputation scores from utilizer feedback comments and an algorithm for mining feedback comments for dimension ratings and weights, coalescing techniques.

M. A. Shafin et al. [4], their goal is to make a system that will analyze the customer's feedback from online shopping and provide a ratio of the positive and negative feedback inscribed in Bangla from the anterior customers utilizing Natural Language Processing (NLP). Amassed over 1000 feedback and comments on the product to conduct the research. We used sentiment analysis along with some relegation algorithms like KNN, Decision Tree, Support Vector Machine (SVM), Arbitrary Forest, and Logistic Regression. With the highest precision of 88.81%, SVM outperformed all the other algorithms.

D.Mali et al. [5], their proposed test model is discussed which will give the direct result of sentiments. It will avail users to accumulate the best information about any particular product they operate, predicated on other customers' reviews, and avail in making a decision about any product.

Aswini U et al. [6], authors have presented a model that includes a multi-dimensional trust model for computing reputation scores from feedback comments, an algorithm for mining feedback comments for dimension rating and weights, combining natural language processing, opinion mining, and topic modeling. Rajkumar Jagdale et al. [7], this analysis is to automatically determine the orientation of a review and contrast. This with the assessment made by the reviewer using Machine Learning Algorithm and NLP is proposed to analyze reviews of Machine learning algorithms such as SVM and NB scoring algorithm (support vector machine).

Thilageswari et al. [8], this paper is to highlight the research design of sentiment analysis and choice of methodology by other researchers in e-commerce customers reviews to guide future development. They proposed 1. Document-level analysis, 2. Sentence level analysis, 3. Aspect level analysis.

Anvar Shathik et al.[9], this literature review explores and discusses the idea of sentiment analysis by undertaking a systematic review and assessment of corporate and community white papers, scientific research articles, journals, and reports.

The primary objective of this paper is to analytically categorize and analyze the prevalent research techniques for sentiment analysis on various applications. This paper includes a research proposal for the e-commerce environment towards sentiment analysis applying ML algorithms.Gurneet Kaur et al. [10], in this they used product review commands and reviews about retailers from flipkart as a data set. It includes spelling correction in review text and then classifying comments by decision tree and NB algorithm. Sentence level sentiment analysis is the main task. It finds details of the comments and has a high degree. Structured reviews are used for testing and training identifying appropriate features and scoring methods from information retrieved for analyzing negative and positive annotations. Then the classifier is used to identify and classify review sentences from the web they have used NB classification. They used the NB algorithm and semantic decision tree to classify the polarity of comments given on websites. First, they use a web crawler to fetch comments. Using WordNet dictionary spelling correction is done. Then stemming is done to remove the stop words. After classifying the overall polarity is calculated using a decision tree.

Tanjim Ul Haque et al. [11], this paper has used the Supervised Learning method on large-scale Amazon datasets to polarize it and get satisfactory accuracy. In this model, they use a manual and active learning approach to label our datasets. Active learning – different classifiers are used to provide accuracy until reaching a satisfactory level, after reaching this they take those labeled datasets and proceed. From the dataset, they extracted features which are then classified by different classifiers.
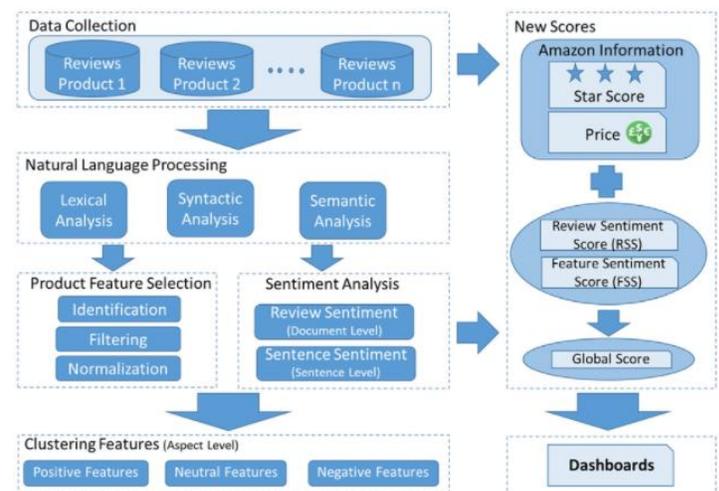
III. PROPOSED METHODOLOGY

Figure. 2 Sentiment Analysis Process Flow using NLP

As shown above, in fig. 2, the process of collecting data from various product reviews of different e-commerce websites is mentioned. In this figure, we can see the complete analysis process with Natural Language Processing (NLP), which is also a base of this project. The complete process is described below:

Process 1: Firstly the data is collected of different products from various websites to start reviewing it.

Process 2: As the data, we have collected is in human language and needs to be converted first in binary codes. So the conversion process is done using NLP.

Process 3: Natural Language Processing shortly known as NLP will help to make natural human language usable to machine easel using NLTK known as Natural Language Toolkit.

Process 4: After the processing system will analyze what kind of ratings and reviews are affecting product selling with sentiments of the words.

Process 5: Analyzing the sentiments of the reviews will help us to understand the product quality and the satisfaction level.
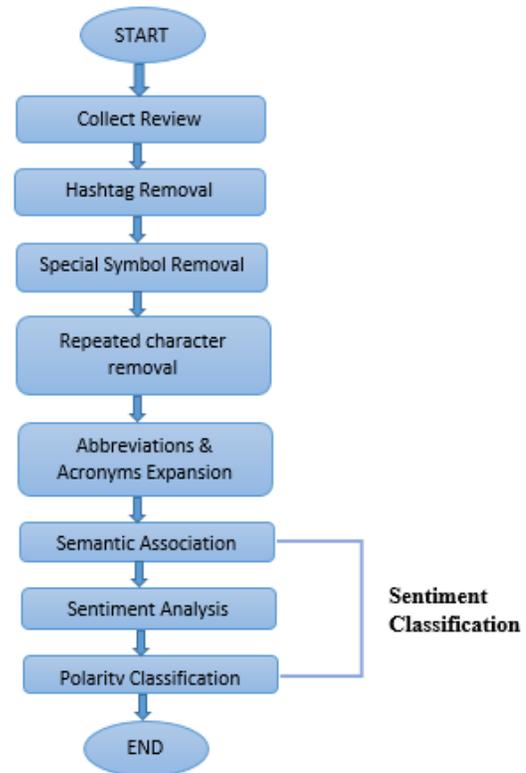

Figure. 3 Sentiment Analysis Flow

The flow defines the complete recognition and the stages of sentiment analysis. With URLs removal, repeat removal, subject classification, subject acceleration.

NLP and NNTK

Natural language processing is like a method to help machines understand the human language with phrases, sentences, short words, etc. The NLTK is a toolkit for a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language.

The above figure shows the processing flow of NLTK which we are using as a base of the project.

Step 1: This will first take raw or unprocessed text from the user or machine.

Step 2: Then the segmentation of the sentence will be performed, this segmentation is referred to as the divination of complete text into layers.

Step 3: Now after the segmentation process the kit will perform tokenization (it is the process where a bigger text is divided into smaller ones).

Step 4: After segmentation and tokenization tagging of the sentence in the form of parts of speech is done to set sentences in various options.

Step 5: In this step entity detection will be performed to check multiple checks of the obtained review.

Step 6: Finally the relation of the obtained review will be compared to the dataset provided to the NLTK to classify the sentiment inaccurate form.

Step 7: And finally the result will be shown on the screen with the sentiment of the review specified. It will also show the accuracy percentage with the analyzed sentiment.

IV.   RESULT
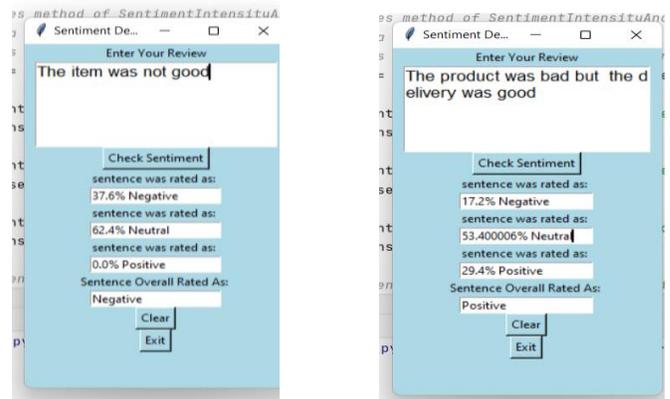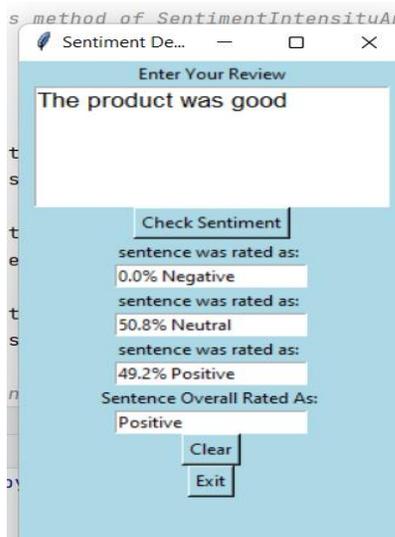
V.   CONCLUSION

In this paper, we discuss the sentiment analysis of the normal words entered by the user. In integration, we do the text processing from data obtained and use Ingenuous Bayes method to soothsayer the class. Afterward, compare with other methods such as NLP. We are relegated by two classes namely positive and negative.

VI.   REFERENCES

[1] Afsal Ali, Jisha P Abraham, Surekha Mariyam Varghese, "A Multi-Dimensional Trust Model and Fuzzy Decision System for ECommerce", International Journal for Scientific Research & Development, Sp. Issue - Data Mining 2015.

[2] Suraj Gund, Akash Gaykwad, Akshada Patil, Payal Nikam, Pankaj Agarkar, "Novel Approach for Mining E-Commerce Feedback Comments using COMM Trust Data Mining Algorithm", International Journal for Scientific Research & Development, Vol. 4, Issue 03, 2016.

[3] Xiuzhen Zhang, Lishan Cui, and Yan Wang, "Computing Multidimensional Trust by Mining E-Commerce Feedback Comments", Journal of LAtex class files, vol. 6, no. 1, January 2017.

[4] M. A. Shafin, M. M. Hasan, M. R. Alam, M. A. Mithu, A. U. Nur and M. O. Faruk, "Product Review Sentiment Analysis by Using NLP and Machine Learning in Bangla Language," 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020, pp. 1-5, doi: 10.1109/ICCIT51783.2020.9392733.

[5] D.Mali, M. Abhyankar, P. Bhavarthi, K. Gaidhar, M. Bangare, "Sentiment Analysis of product reviews for e-commerce recommendation", International Journal of Management and Applied Science, ISSN: 2394-7926 Volume-2, Issue-1, Jan.-2016

[6] Aswini U and Akila Devi T.R, "CommTrust: Computing Multi-Dimensional Trust by Mining E-Commerce Feedback Comments", International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) ISSN: 0976-1353 Volume 22 Issue 2 – MAY 2016.

[7] Brian Keith Norambuena, Exequiel Fuentes Lettura and Claudio Meneses, "Sentiment Analysis and opinion mining applied to scientific paper reviews ", Villegas Intelligent Data Analysis 23 (2019) 191–214 191 DOI 10.3233/IDA-173807

[9] Anvar Shathik, J. &Krishna Prasad, K. (2020). A Literature Review on Application of Sentiment Analysis Using Machine Learning Techniques. International Journal of Applied Engineering and Management Letters (IJAEML), 4(2), 41-77. DOI: http://doi.org/10.5281/zenodo.3977576.

[10] Sentimental Analysis of Flipkart reviewsusingNaïve Bayes and Decision Tree Algorithm, "Sentimental Analysis of Flipkart reviewsusingNaïve Bayes and Decision Tree Algorithm", International Journal of Advanced Research in Computer Engineering &Technology (IJARCET)Volume 5 Issue 1, January2016.

[11] Tanjim Ul Haque, Nudrat Nawal Saber, Faisal Muhammad Shah "Sentiment Analysis on Large Scale Amazon Product Reviews", 2018 IEEE International Conference on Innovative Research and Development (ICIRD)

[12] Springer Nature Singapore Pte Ltd. 2019 P. K. Mallick et al. (eds.), Cognitive Informatics and Soft Computing, Advances in Intelligent Systems and Computing 768, https://doi.org/10.1007/978-981-13-0617-4_61

[13] Zeenia Singla, Sukhchandan Randhawa, Sushma Jain, "Sentiment Analysis of Customer Product Reviews Using Machine Learning", 2017 International Conference on Intelligent Computing and Control (I2C2)

[14] Soumya, S. and K. J. I. E. Pramod (2020). "Sentiment analysis of malayalam tweets using machine learning techniques."

[15] Gamal, D., M. Alfonse, E.-S. M. El-Horbaty and A.-B. M. J. P. C. S. Salem (2019). "Implementation of Machine Learning Algorithms in Arabic Sentiment Analysis Using N-Gram Features." 154: 332-340.

[16] Vanaja, S. and M. Belwal (2018). Aspect-level sentiment analysis on e-commerce data. 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE.

[17] Ezpeleta, E., I. Velez de Mendizabal, J. M. G. Hidalgo and U. J. L. J. o. t. I. Zurutuza (2020). "Novel email spam detection method using sentiment analysis and personality recognition." 28(1): 83-94.

[18] Chakraborty, K., S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag and A. E. J. A. S. C. Hassanien (2020). "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media." 97: 106754.

[19] Ahmad, M., S. Aftab, S. S. Muhammad and S. J. I. J. M. S. E. Ahmad (2017). "Machine learning techniques for sentiment analysis: A review." 8(3): 27.

[20] Arulmurugan, R., K. Sabarmathi and H. J. C. C. Anandakumar (2019). "Classification of sentence level sentiment analysis using cloud machine learning techniques." 22(1): 1199-1209.

[21] Chaturvedi, S., V. Mishra and N. Mishra (2017). Sentiment analysis using machine learning for business intelligence. 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), IEEE.

[22] Hasan, A., S. Moin, A. Karim, S. J. M. Shamshirband and C. Applications (2018). "Machine learning-based sentiment analysis for twitter accounts." 23(1): 11.

[23] Hassan, A. U., J. Hussain, M. Hussain, M. Sadiq and S. Lee (2017). Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. 2017 International Conference on Information and Communication Technology Convergence (ICTC), IEEE.

[24] Kamal, A. and M. Abulaish (2013). Statistical features identification for sentiment analysis using machine learning techniques. 2013 International Symposium on Computational and Business Intelligence, IEEE.

[25] Mukhtar, N., M. A. J. I. J. o. P. R. Khan and A. Intelligence (2018). "Urdu sentiment analysis using supervised machine learning approach." 32(02): 1851001.