



# INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

## Study on Product opinion analysis for customer satisfaction on E-Commerce Websites

Ms. Vijaya Sagvekar  
Research Scholar,

Department of Computer Engineering,  
Pacific Institute of Technology,  
Pacific University, Udaipur (Rajasthan)

Dr. Prashant Sharma

Associate Professor & HOD  
Computer Science & Engineering Department  
Pacific Institute of Technology,  
Pacific University, Udaipur (Rajasthan)

**Abstract**— The E-commerce websites have been emerged in a high range of marketing benefits for the users to publish or share the experience of the received product by posting review that contain useful comments, opinions and feedback on the product. Nowadays, millions of customers gain opportunities to compare similar products in online websites and pick their favorites in digital retailers, such as Amazon.com and Taobao.com. Customer reviews in social media and electronic commerce Websites contain valuable electronic word information of products. Sentiment Analysis is widely applied as voice of customers for applications that target marketing and customer service. Sentiment extractors in their most basic form categorize texts as either having a positive or negative or sometimes neutral sentiment. A common application of sentiment analysis is the automatic determination of whether an online review contains a positive or negative assessment. Hence, in this paper, with the help of the techniques on sentiment analysis, opinionated sentences referring to a specific feature are first identified from product online reviews. We have proposed deep learning method as a classification model for finding the state of review. The results showed recommended site for the customer based on the early reviews, past reviews and answer given to query review for the customer. Moreover, it is observed that the proposed technique can able to reply all the reviews with a better similarity like a human response to the customer.

**Keywords**- E-commerce, opinion analysis, sentiment analysis, deep learning

### I. INTRODUCTION

The marketers and manufacturers have been focused on the market performance from long time. The development of product marketing strategy for managing the product quality is helpful in making a better decision shows improvement in the performance [1]. The manufacturers and marketers continuously collect the product data that helps in analyzing the performance in the market. In the traditional methods, the main data sources for analysis were taken from the manufacturers' internal data, offline customer reviews by the surveys, review forms should be in handwritten format etc [2]. As the technology became advanced, the e-commerce websites enabled to publish the opinion of users on product and the users got a public platform to share their useful comments and opinions towards the product they purchased. According to the survey, 97 % of consumers will be influenced for purchasing the product after going through the reviews posted by other customers. So, a majority of

customers will read online reviews before making an informed purchase decision [3,4].

91 % of people regularly or occasionally read online reviews before purchasing a product and also the early reviews on product before purchasing has high impact on succeeding the product sales [5]. Even though, the early reviewers contribute small proportion of reviews, it is easy to determine the failure or success of new services and new products. Based on the early reviews from the early reviewers helps to adjust marketing strategies in product improvement on designs and helps in succeeding the new product [6]. Based on the early reviews, the companies will recognize the early reviews thereby improve the product designs, marketing strategies that lead to a success on new products. For this reason, early reviewers become the emphasis to monitor and attract at the early promotion stage of a company in the world [7-9]. The early reviews has attracted marketing practitioners extensively analyze the consumer purchase intentions. For example one of the largest e-commerce companies is Amazon in the world, where it provides early reviewer program opportunities for making early reviews that helps the company to acquire those early reviews on products that have few reviews or no reviews. With the help of Amazon shoppers program, will provide information about buying a product by making smarter decision. Based on the above discussions, we can see that early reviewers are extremely important for product marketing [10-14]. The present research will take initiative for studying the behaviour characteristics for the early reviewers and posting it in e-commerce platform such as Amazon, Yelp. The main aim of the research work is to analyze the consumer satisfaction on products bought from E-commerce sites and improved the performance. The overall characteristics of early reviews have to be analyzed from the early reviewers needed to be compared to majority and laggard reviewers. An early reviewer tends to assign a higher average rating score to products and an early reviewer tends to post more helpful reviews. The rating behaviours are characterized that helps to find the scores received from others that helps to determine the correlation of the reviews based on the product popularity [15]. The task is to learn a prediction model which predicts early reviewers given a product. Groups find the positioning ratings for a specific thing with subtle elements that rank last on the clients' opinions, and an ideal features-based positioning method considered.

The findings with the personality variables are related with the theory as follows: higher average rating scores can be considered as the favorable attitude towards the products, and higher helpfulness votes of early reviews given by others can be viewed as a proxy measure of the opinion leadership. Our

analysis also indicates that early reviewers' ratings and their received helpfulness scores are likely to influence product popularity. Herd behavior refers to the fact that individuals are strongly influenced by the decisions of others.

## II. LITERATURE REVIEW

Many researches have been developed for predicting the ranking based on the online reviews. Some of the studies are as follows. In this section, a survey of recent techniques is presented with its advantage and limitations.

Authors	Method	Advantage	Disadvantage
Ahani, A., et al [16]	Self-Organizing Map and Higher Order Singular Value decomposition clustering algorithm	The hybrid algorithm was used for assisting that overcame the data related complications for online reviews and presented spa hotel market segmentation for predicting the travel choice using machine learning algorithms.	The available customer data from TripAdvisor included only the general preferences of spa hotel customers degraded the performance rate
Rita, P et al [17]	Four-dimensions of e-service quality model	The four dimensions of e-service quality were considered such as the impact on customer trust, satisfaction, customer behaviour and building existed literature based on e-service quality during online shopping.	The quality of online stores, in general, was not based on the product segments sold in online stores and the measurements were not applicable for assessing product segments.
Lucini, F.R., et al [18]	Texting Mining approach Latent Dirichlet Allocation (LDA)	The developed model presented a novel framework for customer satisfaction measuring in the airline industry using Latent Dirichlet Allocation (LDA) that detected the popular topic using the natural language processing and machine learning process	The data restricted the diversity of opinions as the proficient in English were in more and likely provided airline experience information.
Bai, T. et al [19]	Margin based Embedding Ranking Model (MERM)	The characterized and predicted the early reviewers for E-commerce sites to present effective product marketing. The developed model used Margin based Embedding Ranking Model (MERM) that predicted the early reviewers in a cold-start setting	The developed model used Margin based Embedding Ranking Model (MERM) that predicted the early reviewers in a cold-start setting.
Zhao, Y.,et al [20]	Technical attributes and sentiment polarity	The review samples were taken from trip advisor that predicted overall customer satisfaction using technical attributes in online and textual reviews of customers.	The textual reviews are influenced by languages and are different cultures needed extension for examining different language reviews.
Liu, Y. et al [21]	Product Competitive and Quality Management and Marketing Strategy	The developed a product competitive advantage analysis for providing an essential basis for quality management and marketing strategy development on social media. The novel method provided an essential basis for managing the marketing strategy and quality using user generated content	The customer comments relied which comes after target product availability of the customers and was not applicable during the design and development of the product.
Jian Jin et al [22]	Product feature extraction and sentiment analysis	The developed model performed opinionated representative for specific product based on the features especially for the competitive products. The sentimental analysis was performed for opinionated sentences that refer specific features for online reviews.	During choosing comparative sentences in the review that were of many for different products resulted lower information comparativeness values.
Sun, Q., et al [23]	Sentiment analysis eWOM	The developed model extracted large volume of online customer reviews and performed sentiment analysis for eWOM products. The developed model used semi-supervised fuzzy product ontology mining algorithm for extraction of features with negative or positive labels.	The developed model required improvement in positive and negative opinion words extraction and also polarity computation.

Thompson, J.J., et al [24]	Lexicon-Based Sentiment Extraction Analysis	The developed model performed sentimental analysis for chat messaging by the player who was involved in video game StarCraft 2. The developed model performed sentimental analysis was applicable for toxicity detection and also identified the players and their messages that are threat for the player.	The human raters disagree not only the sentiment but also disagreed the toxicity that suggested that the task is not straight forward and thus the performance at high rate was impossible
Kumar, S., et al [25]	Sentimental Analysis For The Product Review For EEG Response	The developed model performed multimodal framework for the estimation of product rating on customer product and their brands. The reviews obtained from the global viewers were processed using Natural Language Processing (NLP) technique that computed the score for global rating.	The emotional state needed to be considered for improving the performance.

### III. MOTIVATION FOR STUDY

There are many numbers of studies that investigate the economic outcomes or the drivers of online reviews and, at the same time, propose implications for the design of review systems, such as providing reviewers with a predefined review template. Although a considerable number of review system design features have been proposed over the years, only very few have actually been analyzed. The new online business models and environments have emerged including two-sided platform businesses (e.g., Uber). These enable two-sided reviews and require adapted design features to, for instance, mitigate reciprocity in two-sided review systems.

Lastly, most design features of review systems have been analyzed for stationary devices such as personal computers. However, online reviews are increasingly produced and consumed through mobile devices, which require specific design features. These are the unsolved problems in the ranking analysis for online reviews and many research work based on this problem are still examining.

### IV. PROBLEM DEFINITION

Sentiment analysis performed for the present research classified the text belongs to positive or negative polarity scores alone and the neutral polarity scores were not considered as it is useful in quantifying different affective states of a user[26].

The relationships between customer preferences and design of the product attributes could be highly complex and nonlinear, no theoretical models were developed thus far which are able to model the complex relationships inputting the attribute setting of the new product determine the optimal product attribute settings gave rise to optimization problems.

### V. PROPOSED METHODOLOGY

In our paper, the proposed a application for predict online items in different E-commerce sites like Amazon, Paytm, Flipkart, ShopClues and Snapdeal in reviews of customers. At first, we processed the datasets of different items of particular sources from E-commerce sites. Each item has its own feature sets for categorised depends on a particular features. This features such as significantly positively or negatively influences customer ratings, readability, subjectivity, length–significantly, and sentiment polarity.

The entire surveys of client which refreshed items-based are generated for gathering FCM model[27].

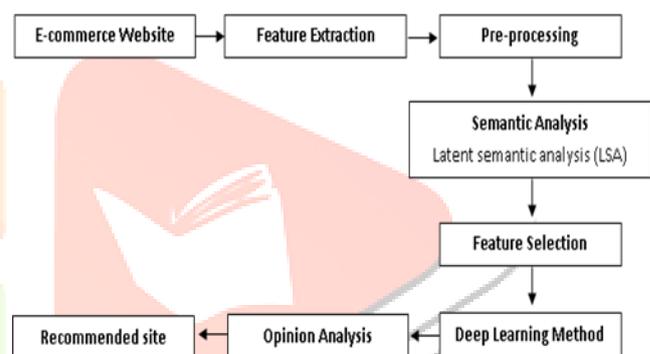


Fig. 1. Flow graph of Proposed Method

#### A. Dataset

The proposed research uses a crude dataset obtained from online site known as E-shopping sites (Amazon, Paytm, Flipkart, ShopClues and Snapdeal). Each and every site is important in introducing a chance to buy the product by great population. The popularity of these sites is identified by various administrations that help to improve the distribution of various products belonging to various categories. Among the E-commerce database, Amazon dataset is used for the present research that has 142.8 million product reviews taken from May 1996 to July 2014. Each of the review from the site consisted of textual comment that was posted by the product user that accompanied to publish time stamp accurate in the study. Usually the reviews are associated scale that is having up to five star of rating that is associated with textual description. A review is associated with a rating score in a five-star scale. Each product is associated with a category label and a textual description.

#### B. Feature Extraction

The reviews from the dataset are extracted using feature extraction technique that gathered item reviews from the E-commerce sites enhanced the quality of review during analysis. The feature extraction includes the items enhancement for performing audit rating of each item. The attributes here refers to the cost, quality, positive survey. The

feature extraction mines the opinion of customer reviews, summarize the reviews, store and produce step preparation.

### C. Pre-processing

Transforming text into something an algorithm can digest is a complicated process. In this section, the steps involved in text processing are as follows.

- (i) Tokenization: The unnecessary tokens are easy to filter, where a document is transformed into paragraphs or sentences into words. In our research work, the online reviews are tokenizing into words.
- (ii) Removal of Unnecessary tags and Punctuation: The next step is to remove punctuation, as the punctuation doesn't do any extra information while treating text data.
- (iii) Removing stop words: Frequent words such as "the", "is", etc. that do not have specific semantic.
- (iv) Lemmatization: Another approach to remove inflection by determining the part of speech and utilizing detailed database of the language.

### D. Semantic Analysis

The semantic analysis is performed for the preprocessed data where, the natural language content reads all the words captures the content by performing real meaning of any text. The text elements are assigned and are assigned based on the logical and grammatical role. The semantic analysis analyzes the surrounding context in the text to accurately disambiguate the exact meaning of words. The relationship between the concepts in the text is also developed to identify the most relevant elements in text and understand the topic discussed. In the semantic analysis, Latent Semantic Analysis (LSA) was used in NLP that analyses the relationship among the set of documents and the terms they contain by producing a set of concepts related to the documents and terms.

### E. Feature Selection

From the extracted concepts, feature selection process is performed for selecting the subset terms that were occurred in training and these selected subsets were treated as features that performed text classification. Firstly, the training is performed and is applied for the classifier decrease the size of the vocabulary effectively. Secondly, the feature selection improves the classification accuracy thereby eliminates the noise features

### F. Deep Learning

The deep learning methods provide an opportunity that faces challenges in NLP problems such as sequence-to-sequence prediction. The developed model performed deep learning methods for learning the features based on the NL which is required by the model specifies the required features and were extracted. The performance of deep learning in natural language processing is based on real results and that the improvements appear to be continuing and perhaps speeding up.

### G. Opinion Analysis

The optimal ranking opinion assesses the significance of each element relatively with respect to the sentiment score that utilized for rank highlights measuring. Based on the opinion analysis, the commentators rated the sites that are having vital data distinguished the untruthful opinions.

### H. Recommended Site

Based on these recommendations, a perspective positioning calculations were performed for ranking the vital angles resulted a viewpoint recurrence that impacted opinions for each perspective for general sentiments.

## VI. RESULTS AND DISCUSSION

With successful implementation and results illustrations, new opinion analysis system is developed for recommending the products more accurately by analyzing the reviews posted for the products by the users. The main motivation behind this experiment is to develop a proper key word extraction method and clustering approach for recommending the products to the customers as negative form, and positive form by using amazon customer review dataset. In this scenario, a key word extraction method (LDA) along with modified GWO algorithm is used for selecting the appropriate key words. The obtained similar key words are clustered using PFCM algorithm. The development automated recommendation system includes numerous advantages; able to identify the fake products, track overall customer satisfaction etc. related to existing classifiers.

Table 1: Comparison of Contingency matrix parameters between proposed systems and different classifiers

	Precision	Recall	F1-Measure	AUC	Accuracy
Random Forest	69.003	73.25	71.132	48.263	73.187
Decision Tree	75.766	75.583	74.618	54.628	75.528
Proposed	76.106	76.781	74.167	57.404	77.236

The proposed system delivered an effective performance by means of quantitative analysis and comparative analysis. From the experimental analysis, the proposed system averagely achieved around 77.236% of classification accuracy, but the existing methodologies attained limited accuracy in amazon customer review dataset. In future work, an effective system is developed in order to further improve the classification accuracy of product recommendation. After implemented our algorithm on multiple Amazon products datasets it has observed final output of proposed system is better as compared to other existing systems.

## VII. CONCLUSION

Now a days the developer community is more and more focusing on the user experience of browsing, because number of users browsing the internet are exponential increased. In this paper, proposed the approach to predict the most effective web-based shopping sites. By client survey determined whether poor or great is the product from business websites. It is more important for any business to know the reviews of clients about its any particular items.

In this paper, to characterize early reviewers on an e-commerce website and to the quantitatively analyze the characteristics of early reviewers and their impact on product popularity will use the Amazon datasets. This proposed study review posting process and develop a deep learning model for the prediction of reviewers. Our model can deal with the cold-start problem by incorporating side information of products. We also mine the opinion of customer reviews, summary of surveys, produce and store for additional step preparation. Based on the opinion analysis, a recommended perspective positioning calculation to rank the vital angles by thinking about both the viewpoint recurrence and the impact of opinions gives the recommended site.

## REFERENCES

- [1] Smith, A.N., Fischer, E. and Yongjian, C., 2012. How does brand-related user-generated content differ across YouTube, Facebook, and Twitter?. *Journal of interactive marketing*, 26(2), pp.102-113.
- [2] Dong, J., Li, H. and Zhang, X., 2014, November. Classification of customer satisfaction attributes: an application of online hotel review analysis. In *Conference on e-Business, e-Services and e-Society* (pp. 238-250). Springer, Berlin, Heidelberg.

- [3] De Giovanni, P. and Zaccour, G., 2014. A two-period game of a closed-loop supply chain. *European Journal of Operational Research*, 232(1), pp.22-40.
- [4] Farhadloo, M., Patterson, R.A. and Rolland, E., 2016. Modeling customer satisfaction from unstructured data using a Bayesian approach. *Decision Support Systems*, 90, pp.1-11.
- [5] Elish, M.O., 2014. A comparative study of fault density prediction in aspect-oriented systems using MLP, RBF, KNN, RT, DENFIS and SVR models. *Artificial Intelligence Review*, 42(4), pp.695-703.
- [6] Clavel, C. and Callejas, Z., 2015. Sentiment analysis: from opinion mining to human-agent interaction. *IEEE Transactions on affective computing*, 7(1), pp.74-93.
- [7] Nahm, Y.E., 2013. A novel approach to prioritize customer requirements in QFD based on customer satisfaction function for customer-oriented product design. *Journal of Mechanical Science and Technology*, 27(12), pp.3765-3777.
- [8] Yang, L., Geng, X. and Liao, H., 2016. A web sentiment analysis method on fuzzy clustering for mobile social media users. *EURASIP Journal on Wireless Communications and Networking*, 2016(1), p.128.
- [9] Yang, J.B., Wang, Y.M., Xu, D.L., Chin, K.S. and Chatton, L., 2012. Belief rule-based methodology for mapping consumer preferences and setting product targets. *Expert Systems with Applications*, 39(5), pp.4749-4759.
- [10] Decker, R. and Trusov, M., 2010. Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing*, 27(4), pp.293-307.
- [11] Dehkharghani, R., Mercan, H., Javeed, A. and Saygin, Y., 2014. Sentimental causal rule discovery from Twitter. *Expert Systems with Applications*, 41(10), pp.4950-4958.
- [12] Chen, A., Lu, Y. and Wang, B., 2017. Customers' purchase decision-making process in social commerce: a social learning perspective. *International Journal of Information Management*, 37(6), pp.627-638.
- [13] Uysal, A.K. and Gunal, S., 2014. The impact of preprocessing on text classification. *Information Processing & Management*, 50(1), pp.104-112.
- [14] Zhang, W., Yoshida, T. and Tang, X., 2011. A comparative study of TF\* IDF, LSI and multi-words for text classification. *Expert Systems with Applications*, 38(3), pp.2758-2765.
- [15] Zhao, T. and Eskenazi, M., 2016. Towards end-to-end learning for dialog state tracking and management using deep reinforcement learning. *arXiv preprint arXiv:1606.02560*.
- [16] Ahani, A., Nilashi, M., Ibrahim, O., Sanzogni, L. and Weaven, S., 2019. Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. *International Journal of Hospitality Management*, 80, pp.52-77.
- [17] Rita, P., Oliveira, T. and Farisa, A., 2019. The impact of e-service quality and customer satisfaction on customer behavior in online shopping. *Heliyon*, 5(10), p.e02690.
- [18] Lucini, F.R., Tonetto, L.M., Fogliatto, F.S. and Anzanello, M.J., 2020. Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. *Journal of Air Transport Management*, 83, p.101760.
- [19] Bai, T., Zhao, W.X., He, Y., Nie, J.Y. and Wen, J.R., 2018. Characterizing and predicting early reviewers for effective product marketing on e-commerce websites. *IEEE Transactions on Knowledge and Data Engineering*, 30(12), pp.2271-2284.
- [20] Zhao, Y., Xu, X. and Wang, M., 2019. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, 76, pp.111-121.
- [21] Liu, Y., Jiang, C. and Zhao, H., 2019. Assessing product competitive advantages from the perspective of customers by mining user-generated content on social media. *Decision Support Systems*, 123, p.113079.
- [22] Jin, J., Ji, P. and Gu, R., 2016. Identifying comparative customer requirements from product online reviews for competitor analysis. *Engineering Applications of Artificial Intelligence*, 49, pp.61-73.
- [23] Sun, Q., Niu, J., Yao, Z. and Yan, H., 2019. Exploring eWOM in online customer reviews: Sentiment analysis at a fine-grained level. *Engineering Applications of Artificial Intelligence*, 81, pp.68-78.
- [24] Thompson, J.J., Leung, B.H., Blair, M.R. and Taboada, M., 2017. Sentiment analysis of player chat messaging in the video game StarCraft 2: Extending a lexicon-based model. *Knowledge-Based Systems*, 137, pp.149-162.
- [25] Kumar, S., Yadava, M. and Roy, P.P., 2019. Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. *Information Fusion*, 52, pp.41-52.
- [26] 'Pradip Suresh Mane, Ashok Kumar Jetawat, Pravin Jagannath Nikumbh', 2020, "Web Page Recommendation using Random Forest with Fire Fly Algorithm in Web Mining", *International Journal of Engineering and Advanced Technology (IJEAT)*, Vol. 9, Issue-3, February 2020, pp. 499-505.
- [27] Kolhe L.N., Khairnar V., Jetawat A.K. :Prediction-Based Parallel Clustering Algorithm for M-Commerce. In: Fong S., Akashe S., Mahalle P. (eds) *Information and Communication Technology for Competitive Strategies. Lecture Notes in Networks and Systems*, vol 40. Springer, Singapore, pp. 31-39.

## AUTHORS PROFILE

**Ms. Vijaya Sagvekar** is currently pursuing PhD (Computer Engineering) and working as an Assistant Professor in Information technology department at Vasantdada Patil Prathishathan's college of engineering, Mumbai. She has teaching, research experience of 10 years. Database management systems, Software engineering, Web Mining, Data Optimization, Network security are just some of his areas of interests.

**Dr. Prashant Sharma** is having 8 years of teaching experience and working as an Associate Professor in the Computer Science and Engineering department of Pacific University, Udaipur (Rajasthan). He has completed Ph.D. in September 2018, MTech (CSE) in 2012 and BE (IT) in 2007. His area of interest and research area is Nature-Inspired Algorithms and Machine learning.

