

**SENTIMENTAL ANALYSIS FOR USER REVIEWS USING DUAL SENTIMENT
ANALYSIS TECHNIQUES**Jayesh Patil¹, Gaurav Wagh², Parth Dhonde³, Varsha Dange⁴¹B.E Student, Dept. of Computer, Dhole Patil College of Engineering, Pune, Maharashtra, India²B.E Student, Dept. of Computer, Dhole Patil College of Engineering, Pune, Maharashtra, India³B.E Student, Dept. of Computer, Dhole Patil College of Engineering, Pune, Maharashtra, India⁴Asst.Professor, Dept. of Computer, Dhole Patil College of Engineering, Pune, Maharashtra, India

Abstract-Sentiment analysis or opinion mining aims to detect subjective information such as opinions, attitudes, and feelings expressed in text by the people. Opinion mining would process a set of search results for a given item, generating a list of product attributes like its quality, features; and aggregating opinions about each of them like poor, mixed, good. Then it will identify the unique properties of this problem and develop a method for automatically distinguishing between positive, negative and neutral reviews. Sentiment analysis gathers the public opinions and reviews. Dual sentiment analysis considers two sides of one review. It is used to analyze the user reviews for the product. Sentiment analysis is used for tracking the mood of the public about a particular product or topic using natural language processing. Sentiment analysis is also called as opinion mining it involves in building a system which collect and examine opinions about the product made in comments, reviews, or tweets. With the vast growth of the social media contents and online shopping sites on the Internet in the past few years, people now express their opinion on almost anything. Searching the user opinions on sites and monitoring them on the web is difficult task. Thus there is a need for automatic opinion Sentiment Analysis is the area of study that analyzes customer feedback, opinions, sentiments, evaluations, attitudes, and from written language. It is one of the most useful research areas in Natural Language Processing and is also widely studied in data mining. The importance of sentiment analysis coincides with the growth of social media such as reviews, discussion forums, blogs, and social networks. Additionally this system adds the hit count and the buy count features so as to find the number of users who have visited the site and purchased the product. The objective of this feature is to know the amount of people who have browsed the product and have not purchased, this will register in the hit count class. If a particular person bought the product this will make sure that it gets registered in the buy count. The main idea is to let the potential customer know the amount of people who have just browsed the product and the people who have scrolled through the product. This will give an in detail depth knowledge and the popularity of these product.

Keywords:-BOW-Bag Of Words, DSA-Dual Sentiment Analysis, Naïve Bayes, Opinion mining, SVM-support vector machine.

I. INTRODUCTION

Sentiment analysis is the current research topic in the field of data mining. It is difficult to extract Opinions and sentiments from natural language. Sentiments are extracted from user comments, reviews and feedbacks. "What other people think" has always been an important piece of information while taking any decision.

Few years ago, if an individual need to take a decision he used to ask for opinions from friends and family. If organization wanted to find opinions of the general public about its products and services, it conducted surveys. With the increase in social media content on the internet in the past few years, the world has been transformed. E Commerce sites and social media are some of the platforms on which people can now express their views on almost anything in discussion Now days, before planning to buy a product, everyone what to know its reviews. Sentiment analysis is the best tool for finding whether the review is positive or negative. It helps people to find good quality product. It also helps companies by providing customers feeling related to their product. It also helps to analyze public sentiments related to political issues or political candidates. The main aim of the system is to analyze sentiments for E-shopping company services. The reviews are classified into three categories positive, negative and neutral. These results can guide us to select particular product, based on maximum number of positive reviews for products.

Sentiment classification is a basic task in sentiment analysis, to classify the sentiment (e.g., positive or negative) of a given text. the bag-of-words (BOW) model is typically used for sentiment classification. In the BOW model, a Review text is represented by a vector of independent words to train a sentiment classifier statistical machine learning algorithms (such as

naïve Bayes, maximum entropy classifier, and support vector machines) are then employed. The BOW model is very simple and quite efficient.

II. RELATED WORK

Abinaya.R et al.[1] have proposed a method to classify the sentiments with the help of dictionary into different categories such as positive, negative and neutral. It uses support vector machine algorithm to improve quality of classification of reviews. It also showed the graphical representation of review category. BOW is not used as it has sum deficiency to handle polarity shifting problem.

Bhagyashri Jadhav et al.[2] has discussed various types of sentiment analysis techniques such as supervised and unsupervised learning. Steps of sentiment analysis are also discussed which are as follows, text extraction, text refinement and text classification. Supervised approach is also called as machine learning. Unsupervised approach is a type of natural language processing.

V.S.Swarnalekha et al.[3] have discussed the polarity shifting issue in sentiment classification. Bag Of Words is used to classify the sentiments into different categories. Naive Bayes and Support vector machine algorithms are easily adaptable to the dual training architecture discuss in this paper. Lexican based approach is confidential for dictionary and corpus based method.

Shulong Tan et al.[4] have discussed the sentiment analysis on twitter and how it will help to make the decisions. The product seller can know about the product quality by the sentiments of people and he can decide whether the product should be produce in large quantity or not. It also helps to improve the product quality. The buyers can also decide whether to buy the product or not with the help of sentiments. Decision making is made easier by knowing the exact reason behind the variations of sentiments on twitter using RCBLDA model.

Dr. U Ravi Babu[5] has proposed the use of “Sentiwordnet dictionary” to find the scores of each words. The sentiments of people on online shopping sites are gathered. The services of online shopping sites are compared to find out the best online shopping site. For this the reviews related to services from five large datasets of the online shopping sites is used. The sentiwordnet dictionary gives the best performance and is more reliable. The preprocessing of data affects the quality of detected sentiments.

Rui Xia et al.[6]This paper discussed how to tackle the polarity shifting issue in sentiment analysis. Polarity shifting makes the negation of statement. Dual Sentiment analysis model is used to solve the polarity shifting issue. In Bag of Words the two opposite sentiment texts are considered to be same which causes polarity shifting.

III. PROPOSED SYSTEM

In previous papers the sentiments were only classified and graph was produced. This paper proposed a system in which buy count and hit count will be shown. Buy count is the total number of people how buy this product and hit count is the total number of people how viewed this product. Propose a simple yet efficient model, called dual sentiment analysis (DSA), to address the polarity shift problem in sentiment classification Figure 1. Show a System of proposed architecture. A data expansion technique by creating sentiment-reversed reviews. The original and reversed reviews are established in a one-to-one correspondence & dual training algorithm and a dual prediction algorithm correspondingly, to make use of the original and reversed specimen in pairs for training a statistical classifier and make predictions. DSA structure from polarity (positive-negative) classification to 3-class (positive, negative, neutral) sentiment classification.

3.1 Dual Training.

The original training specimens are reversed to their opposites. Indicate to them as “original training set” and “reversed training set. In our data expansion technique, there is a one-to one correspondence among the original and reversed reviews. The classifier is trained by maximizing a combination of the likelihoods of the original and reversed training samples. This process is called dual training. Note that our method can be easily adapted to the other classifiers such as naïve Bayes and SVMs.

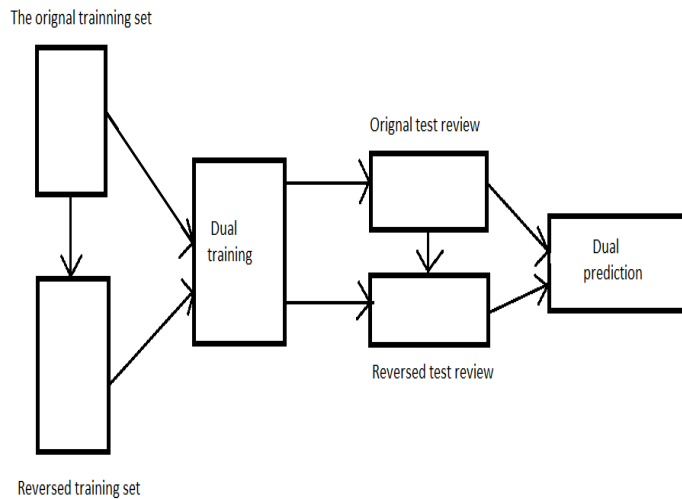


Figure 1.*The process of Dual Sentiment Analysis*

III. Dual Prediction.

Dual prediction works in addressing the polarity shift problem. This time we think “I don’t like this book. It is boring” is an original test review, and “I like this book. It is interesting” is the reversed test review. Accordingly, it is very likely that the original test review will be misclassified as Positive. While in Dual Prediction, due to the removal of negation in the reversed review, “like” this time the plays a positive role. Therefore, the probability that the reversed review being classified into Positive must be high. In Dual Prediction, a weighted combination of two component predictions is used as the dual prediction output.

IV. MATHEMATICAL MODEL

- Let S be the system that describes the product review extraction, pre-processing, sentiment labeling, sentiment analysis.
- $S = \{Pr, Pt, S1\}$
- Pr =Product Review
- $S1 = \{Pv, Nv, Ne\}$
- $Pv = \{P1, P2, \dots, Pn\}$ =Positive class
- $Nv = \{N1, N2, \dots, Nn\}$ = Negative class
- $Ne = \{Ne1, Ne2, \dots, Nen\}$ = Neutral class

Where,

- S =Sentiment analysis system.
- Pt = Preprocessing of product review(Slang word translation, Non-English word and URL removal).
- $S1$ = Classify review in three category(Positive, Negative, Neutral).
- $P1, P2 \dots Pn$ = Positive product review class.
- $N1, N2 \dots Nn$ = Negative product review class.
- $Ne1, Ne2 \dots Nen$ = Neutral product review class.

V. ALGORITHM

The Naive Bayesian classifier is based on Bayes’ theorem with independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Despite its simplicity, the Naive Bayesian classifier often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assume that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability
Posterior Probability
Predictor Prior Probability

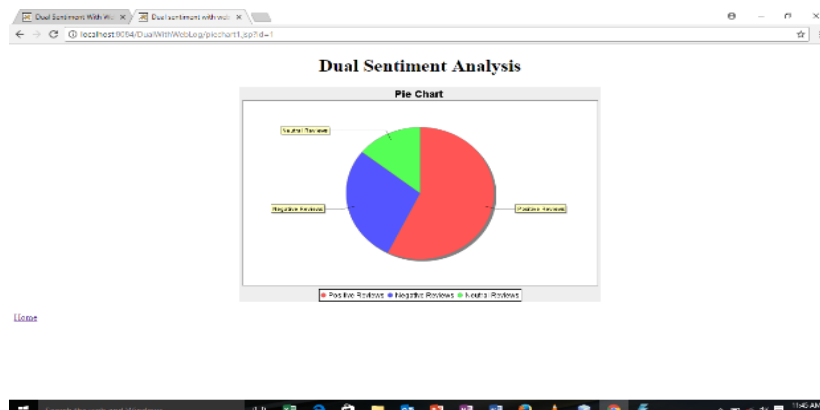
$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Above,

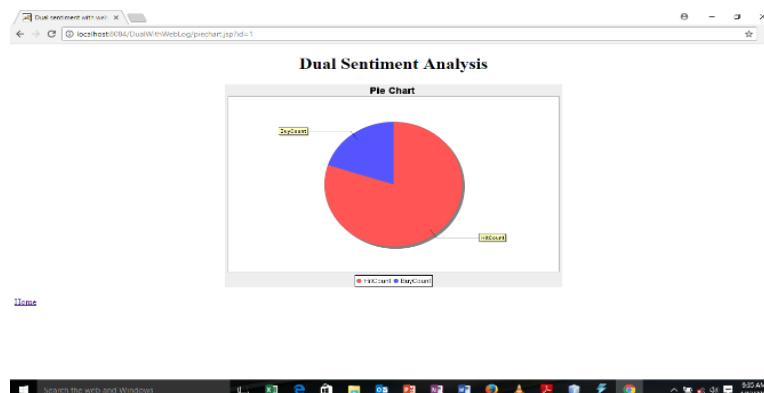
- $P(c|x)$ is the posterior probability of class (c , target) given predictor (x , attributes).
- $P(c)$ is the prior probability of class.
- $P(x|c)$ is the likelihood which is the probability of predictor given class.
- $P(x)$ is the prior probability of predictor.

VI. RESULT

Graph Of Sentiment Classification:-



Graph of Buy Count and Hit Count:-



In the existing system buy count and hit count are not shown, whereas in this system the graphical representation of buy count and hit count is shown. Also the graphical representation of sentiment category such as positive, negative and neutral is represented as show in figure.

VII. CONCLUSION

Sentiment analysis is useful for anyone who is going to make a decision. Sentiment analysis is helpful in different field for identifying and expressing sentiment. It is helpful for everyone who want to buy a product and they can decide which product is best. Sentiment analysis helps Enterprises to know what customers think about their products. Therefore companies can take decisions about their products and its qualities based on customer's feedback. Thus companies can modify their products features and introduce new products according to customers' opinions in a better and faster way.

REFERENCES

- [1] Abinaya.R, Aishwaryaa.P and Baavana.S "Automatic Sentiment Analysis Of User Reviews," 2016 IEEE International Conference On Technological Innovations in ICT For Agriculture and Rural Development.
- [2] Bhagyashri Jadhav and Manjushri Mahajan "Review of Dual Sentiment Analysis," International Journal Of Science and Research(IJSR), 2015.
- [3] V.S.Swarnalekha and V.M.Suresh "Sentiment Analysis For Two Sides Of Review Using Dual Prediction," International Journal Of Technology and Engineering Systems, Vol.8 No.1-Jan-March 2016.
- [4] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan and Xifeng Yan "Interpreting the Public Sentiment Variations on Twitter," IEEE Transactions on Knowledge and Data Engineering, Volume 26 Issue 5, May 2014.
- [5] Dr. U Ravi Babu "Sentiment Analysis of Reviews for E-Shopping Websites" International Journal Of Engineering And Computer Science, Volume 6 Issue 1 1 Jan. 2017.
- [6] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li "Dual Sentiment Analysis: Considering Two Sides of One Review," IEEE Trans. Know.Data Eng., vol. 27, no. 8, Aug. 2015.
- [7] V. Hatzivassiloglou and J.M. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," Proc 18th Conf. Computational Linguistics, pp. 299-305, 2000.
- [8] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics, 2004.
- [9] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment Classification Using Machine Learning Techniques," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 79-86, 2002.
- [10] Q. Su, X. Xu, H. Guo, Z. Guo, X. Wu, X. Zhang, B. Swen, and Z. Su, "Hidden Sentiment Association in Chinese Web Opinion Mining," Proc. 17th Int'l Conf. World Wide Web, pp. 959-968, 2008.
- [11] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, and W. Zuo, "Sentiment Analysis of Chinese Documents: From Sentence to Document Level," J. Am. Soc. Information Science and Technology, vol. 60, no. 12, pp. 2474-2487, Dec. 2009.
- [12] P.D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," Proc. 40th Ann. Meeting on Assoc. for Computational Linguistics, pp. 417- 424, 2002.
- [13] Dhiraj Gurkhe, Niraj Pal, Rishit Bhatia, "Effective Sentiment Analysis of Social Media Datasets using Naive Bayesian Classification", International Journal of Computer Applications (0975 8887), Volume 99 - No. 13, August 2014.
- [14] Riya Suchdev, Pallavi Kotkar, Rahul Ravindran, "Twitter Sentiment Analysis using Machine Learning and Knowledge-based Approach", International Journal of Computer Applications (0975 – 8887), Volume 103 – No.4, October 2014.
- [15] 14 Eman M.G. Younis, "Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study", International Journal of Computer Applications (0975 – 8887), Volume 112 – No. 5, February 2015.