Sentiment Analysis of Product Review

Chandu Vaidya, Rutika Janbandhu, Sampada Dahikar, Shital Lichade, Ekta Khode

Rajiv Gandhi College of Engineering and Research, Wanadongri, Nagpur, Maharashtra, India

A R T I C L E I N F O

A B S T R A C T

Sentiment analysis is defined as the process of mining of data, view, review or sentence to predict the emotion of the sentence through natural language processing (NLP). The sentiment analysis involves classification of text into three phase "Positive", "Negative" or "Neutral". It analyses the data and labels the 'better' and 'worse' sentiment as positive and negative respectively. Using social media, e-commerce website, movies reviews such as Facebook, twitter, Amazon, Flipkart etc. user share their views, feelings in a convenient way. To analysis of such huge data automatically, the field of sentiment analysis has turn up. The main aim of sentiment analysis is to identifying polarity of the data in the Web and classifying them. Therefore, to find polarity or sentiment of, user or customer there is a demand for automated data analysis techniques. In this paper, a detailed survey of different techniques or approach is used in sentiment analysis and a new technique which is proposed in this paper.

Keywords : Sentiment Analysis, Naive Bayes, Mining, Support Vector Machine, Supervised approach, Unsupervised approach, Polarity, Semantic

I. INTRODUCTION

To work with raw data proper methods required. Most of the methods either focus on verbs, nouns, adverbs or adjectives. But no work has focused on all the possible combinations of adverbs, adjectives and verbs. This paper presents the theoretical analysis of some well-known methods or proposal of Sentiment Analysis [1]. The sentiment analysis involves classification of text into three phase "Positive", "Negative" or "Neutral". Artificial intelligence is the simulation of human intelligence processes by machines, especially computer systems is Artificial intelligence[2]. NLP is the computer program which understand human language as it is spoken and written – referred to as natural language[3]. Support Vector Machine or SVM is Supervised Learning algorithms, which is used for Classification as well as Regression problems[4]. However, primarily, it is used for Classification problems in Machine Learning. The following combinations are taken into for analysis, adverbs-adjectives, adverbs-verbs, adjectives-verbs and adverbs-adjectivesverbs along with adverbs, adjectives and verbs. Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural

networks attempt to simulate the behavior of the human brain—albeit far from matching its ability—allowing it to "learn" from large amounts of data The Standard classifier like Naive Bayes (NB)[5], Linear Model and Decision Tree are used to deduct result and for analysis.

II. Background

A. Product Reviews

User reviews are reviews from product consumers, which written with natural language, and contain various types of structure for expressing their opinion towards product . They often use formal natural language, present well organised structure, and also included a review summary such as number of stars or score indicating satisfaction. This resulted reviews from Steam is more consistent because a product can be developed form or than one Ecommerce sites. User reviews in steam consist of

1) Product name
2) Product reviewer information which are username, address where the product will deliver, product reviewed , and reviewed date 3) Review text
4) Number of helpful and not helpful rating from other users and
5) Product review comments from other users.

B. Text Mining

The concept of text Mining [6] is to find valuable patterns and relations from a set of textual data which is usually gathered from a collection of unstructured documents. Text mining has been applied into various user and one of them is text classification.

There are 2 approaches for text classification : supervised approach and unsupervised approach.

Supervised approach[7] requires training data to accurately classify text. Unsupervised approach[8] does not use training data and one of method in this approach is compiling set of keywords into text corpus to classify text by keywords that appeared on text.

III. Literature Survey

The [1] Mohammed H. Abd El-Jawad and et. al has proposed in the year 2020 about rise of social media, such as Twitter platform compares the performance of different machine learning and deep learning algorithms, in addition to introducing a new hybrid system that uses text mining and neural networks for sentiment classification and got result maximum accuracy rate of 83.7%. The [2] Brandon Joyce and et. al has proposed in year 2016 about manually labeled tweets as well as automatically labeled tweets based on hashtag content/topic using a lexicon and Naive Bayes Machine Learning Algorithm & got result 48.8% accuracy. The [3] Mochamad Mashuri & et. al has proposed in year 2019 about sentiment analysis is determining the polarity or type of opinion in a predetermined text or subject. NLP (Natural Language Processing) technique is used to support the beginning of a text and got result 68.3%.

The [4] Adyan Marendra Ramadhani has proposed in year 2017 about twitter sentiment analysis used deep learning algorithm & got result 75% accuracy. The [5] Cem Baydogan and et. al has proposed in year 2018 about Konstanz Information Miner (KNIME), which is a powerful data mining tool with its richest features and many visualization tools, was used on twitter data using machine learning algorithms on Twitter data & got result 75.2% accuracy. The [6] Sharmishta Desai and et. al has proposes in year 2015 about Analysis of social media data can be used for many purposes like friend recommendation, product or service recommendation etc. using decision tree algorithm & got result 90.34% accuracy.

The [7] Nicholas Mastronarde and et. al has proposed in year 2010 about allowing the multimedia system layers to learn, through repeated interactions with each other, to autonomously optimize the system's long-term performance at runtime using reinforcement learning algorithms & got result 92 %

accuracy. The [8] Tejas Mankar & et. al has proposed in year 2019 about stock investments using Machine learning and artificial intelligence techniques and got result 46% accuracy. The [9] Gergo Hajdu and et.al has proposed in year 2007 about Facebook friend request is authentic or not using artificial neutral network and got result

43.8% accuracy.

The [10] Meredita Susanty and et. al has proposed in year 2019 about develop offensive content detection for posts written in Bahasa uses an artificial neural network model & got result 99.18% excellent accuracy. The [11] Abdhullah Al Manun has purposed in year 2018 about social media bullying detection using machine learning algorithm & got result 97%. The [12] Gergo Hajdu and et.al has proposed in year 2007 about Facebook friend request is authentic or not using artificial neutral network and got result 43.8% accuracy.

The [13] Adyan Marendra Ramadhani has proposed in year 2017 about twitter sentiment analysis used deep learning algorithm & got result 75% accuracy. The [14] Sharmishta Desai and et. al has proposes in year 2015 about Analysis of social media data can be used for many purposes like friend recommendation, product or service recommendation etc. using decision tree algorithm & got result 90.34% accuracy The [15] Tejas Mankar & et. al has proposed in year 2019 about stock investments using Machine learning and artificial intelligence techniques and got result 46% accuracy

IV. Problem Statement

Sentiment and opinion have an important characteristic. First of all, different people may have different experiences and thus different opinions. An application that collects reviews from the users about a certain product and analyses them. It would segregate the reviews into positive and negative reviews. The application further provides the pros and cons of the individual feature of the product. In our project, we have sites such as Amazon.com, flipkart.com,

rediff.com & many more with help of these we find the customer review. Product review is used on shopping sites to give customers an opportunity to comment on product they have purchased, right on the product page.

V. Aim & Objectives

AIM: Illustrate the Customer review with the help of sentiment analysis.

OBJECTIVE: Social media refers to the means of interactions among people in which they create, share, or exchange information and ideas in virtual communities and networks. Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.To give customers an opportunity to rate and comment on product they have purchased, right on the product page is the product review.
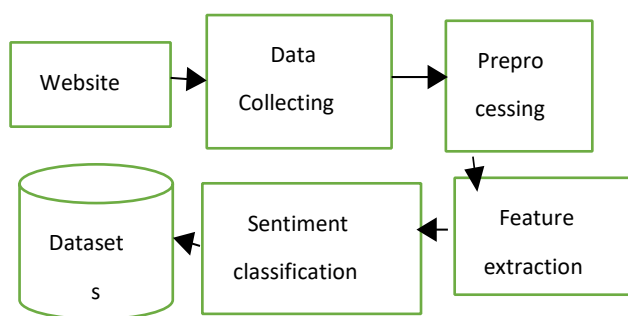
VI.Proposed Approach

Fig no.1 Flow Diagram

Explanation of diagram is as follow: 1) **Data Collecting:** The first step in any data mining application is gaining data to work . The first is explicit knowledge acquisition, where the user is required to manually enter information about themselves. The second is implicit, or passive data acquisition, where data is collected as a user .
2) **Preprocessing:** The text document preprocessed and stored in an appropriate data structure for further

processing. 3) **Feature Extraction:** This is an important step in product opinion mining, that can be classified into four categories: machine learning, ontology, lexicon-based and dependency-relation-based approaches. We wanted to classify the data into positive, negative and neutral categories.

4) **Sentiment classification:** is a method to distinguish opinions of the sentence. Sentiment of opinions often classifies as positive or negative opinion. 5)**Datasets:** Flipkart does not influence ratings and reviews on the platform. Every detail from the product images and specifications, to description and details is updated on the site. If you, as a customer, write a review about a smartphone that you recently purchased on Flipchart will have a large influence on the decision of other shoppers who are looking to buy the same brand or model of smart phones.

VII.   PROPOSED ALGORITHM

This section illustrates the proposed algorithm for sentiment analysis. This proposed algorithm is divided into three phases

The Naiye bayes algorithm :

Naïve Bayes algorithm is a supervised learning algorithm, which based on bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a highdimensional training dataset. Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in quick predictions.

Firstly, the the data is trained it means the data model is used as for assumption and then this trained data is being tested.

For solving the probability, we use the formula :

$P(A/B) = P(B/A) *P(A)/P(B)$

From this mathematical formula we calculate the probability means the reviews. It means this algorithm give the result in probability of that review is good, bad or neutral.

VIII.   IMPLEMENTATION

A product review is an opinion about certain product features, performance, quality, and overall value. The purpose of a product review is to provide potential to buyers with a detailed and honest informed decision before purchasing.

Reviews are social proof to build the trust and loyalty. For the project we use Navies

Bayes algorithm. Naive Bayes algorithm is a supervised learning algorithm, which based on bayes theorem and used for solving classification problems. It is mainly used in *text classification* that includes a highdimensional training dataset. Naive Bayes Classifier is one of the simple and most effective Classification algorithms which helps in quick predictions.

Firstly, convert the given dataset into tables. Generate the probabilities of given features. Now, use Bayes theorem to calculate the probability.

The grammar use is Natural language processing (NLP). NLP is the computer program which understand human language as it is spoken and written – referred to as natural language. It is a component of artificial intelligence. This grammar perform and goes under five phases : lexical analysis, Syntactic analysis, Semantic analysis, discourse analysis, Pragmatic analysis.

As the final result, it produces 97 % positive accuracy, 7% negative.

IX.   CONCLUSION

Here we conclude that using the grammar we produce the reviews from the naïve bayes as from sentiment analysis.

X. FUTURE WORK

It's been an interesting field of study. The ability to understand sarcasm, hyperbole, positive feelings, or negative feelings has been difficult, for machines that lack feelings. There is too much potential in machine learning.

XI. REFERENCES

[1]. Mrs. Pranjal S. Bogawar Assistant Professor, Department of Information Technology, Priyadarshini College of Engineering, R.T.M. Nagpur University, IEEE.

[2]. Layton, R., Watters, P., Dazeley, R. 2010. Authorship Attribution for Twitter in 140 characters or less. Second cybercrime and trustworthy computing workshop, IEEE. (July 19-20, 2010) 1-8. DOI=10.1109/CTC.2010.17

[3]. Vel, O. De, Corney, M., Anderson, A. and Mohay, G. 2002. Language and Gender Author Cohort Analysis of E-mail for Computer Forensics. Digital Forensics Research Workshop (2002)1-16. DOI=10.1.1.19.8168

[4]. Vel, O. De, Corney, M., Anderson, A. and Mohay, G. 2001. Mining E-mail content for author identification Forensic. ACM Sigmod record.3,4 (New York, NY, USA, 2001) 5564. DOI=10.1.1.20.697

[5]. Al Fe'ar, N., Al Turki, E., Al Zaid, A., Al Duwais, M., Al Sheddi, M., Al khamees, N., & Al Drees, N. 2008. Eclassifier: A Bi-Lingual Email Classification System. Information Technology ITSim 2008 IEEE (2008) 2:1-4. DOI= 10.1109/ITSIM.2008.4631707

[6]. Samaneh Moghaddam and Martin Ester, "Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews", Proceedings of 19th ACM International Conference on Information and Knowledge Management, pp. 18251828, 2010.

[7]. Aurangzeb Khan, Baharum Baharudin and Khairullah Khan, "Sentiment Classification from Online Customer Reviews using Lexical Contextual Sentence Structure", Proceedings of International Conference on Software Engineering and Computer Systems, pp. 317-331, 2011.

[8]. M. Hu and B. Liu, "Mining and Summarizing Customer Reviews", Proceedings of 10th ACM International Conference on Knowledge Discovery and Data Mining, pp. 166-177, 2005.

[9]. A.M. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews", Proceedings of International Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 339-346, 2005.

[10]. G. Vinodhini and R.M. Chandrasekaram, "Sentiment Analysis and opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 6, pp. 28-35, 2012.

[11]. A. Collomb, C. Costea, D. Joyeux, O. Hasan and L. Brunie, "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation", Available at: https://liris.cnrs.fr/Documents/Liris6508.pdf.

[12]. Mika V. Mantyla, Daniel Graziotin and Miikka Kuutila, "The Evolution of Sentiment Analysis-A Review of Research Topics", Computer Science Review, Vol. 27, No. 1, pp. 16-32, 2018.

[13]. F. Benamara, C. Cesarano and D. Reforgiato, "Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone", Proceedings of International Conference on Weblogs and social media, pp. 1-7, 2006.

[14]. R.A. Hummel and S.W. Zucker, "On the Foundations of Relaxation Labeling Processes", Proceedings of International Conference on Computer Vision: Issues, Problems, Principles, and Paradigms, pp. 585-605, 1987.

[15]. Samaneh Moghaddam and Martin Ester, "ILDA: Interdependent LDA Model for Learning Latent Aspects and their Ratings from Online Product Reviews", Proceedings of 34th International ACM Conference on Research and Development in Information Retrieval, pp. 665-674, 2011

[16]. Jorge Carrillo De Albornoz, Laura Plaza, Pablo Gervas and Alberto Diaz, "A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating", Proceedings of International Conference on Advances in Information Retrieval, pp. 55-66, 2011.

[17]. Sentiment Analysis, Available at: https://insightsatlas.com/slysis/

[18]. Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts, "Learning Word Vectors for Sentiment Analysis", Proceedings of 49th Annual Meeting of the Association for Computational Linguistics, pp. 1-7, 2011. [19] Understanding Sentiment Analysis: What It Is and Why It's Used Understanding Sentiment Analysis: What It Is and Why It's Used, Available at: https://www.brandwatch.com/blog/understandingsentiment-analysis/

[19]. Sentiment Analysis Explained, Available at: https://www.lexalytics.com/technology/sentiment-analysis

Cite this article as :

Chandu Vaidya, Rutika Janbandhu, Sampada Dahikar, Shital Lichade, Ekta Khode, "Sentiment Analysis of Product Review ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.10-15, May-June-2023.