Product Reviews on Opinion Mining using NLP Techniques

Ms. Lina Dinesh, Assistant Professor, Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore-641 202, Tamil Nadu, India.

linadinesh@sece.ac.in

Ms. Kousalva A, Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore-641 202, Tamil Nadu,

India.

kousaly a.a2018cse@sece.ac.in

Ms. Kaviya G, Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore-641 202, Tamil Nadu, kaviya.g2018cse@sece.ac.in

Ms. Lavanya N, Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore-641 202, Tamil Nadu, India.

lavany a.n2018cse@sece.ac.in

Ms. Kavya N C, Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore-641 202, Tamil Nadu,

kavya.n.c2018cse@sece.ac.in

Abstract-- Opinion mining is a natural language processing (NLP) based automated text analysis to identify emotions and opinions (positive and negative comments) present in a text. Opinion mining is the technique used for automatically gathering knowledge from different user opinions on a certain topic or situation. The amount of data available on the internet is growing at a high rate. Every day, social media generates a massive quantity of data, such as reviews, comments, and customer feedback. This huge volume of user-generated data is unintelligible unless specific mining techniques are applied to it. Since there are so many fake reviews, there is a need for the opinion mining approach to integrate a spam detection module to deliver a legitimate opinion. People nowadays rely on social media opinions to decide whichever products or services to purchase. Since there are numerous bogus or fraudulent reviews issued by organizations or individuals for a number of reasons, spam detection will emerge as a complex and time-consuming task. The proposed system includes ontology, geolocation and IP address monitoring, a lexicon of spam phrases using naive bayes, comment detection, and account tracking.

Keywords - Opinion Mining, NLP, Online Reviews, POS tag, Stop Words, Stemming, Dataset, Geolocation and IP Address Tracking.

I. INTRODUCTION

Opinion mining makes use of text analysis to gain a better understanding of the factors that influence public opinion. In recent times, opinion mining has preceded sentiment analysis [1]. For example, sentiment analysis examines how people feel about a certain matter (positive or negative), as contrasted to opinion mining, which investigates why people feel the way they do. The complier must understand and develop these components while considering the following requirements.

Based on the information gathered from recent reviews, this article presents an opinion mining framework. It is a research field at the crossroads of information retrieval. Moreover, natural language processing has some things in common with other fields such as text mining and information extraction [23]. To assess and summarize the opinions expressed automatically by the computer, some opinion data is required. It helps to determine a person's

positive, negative and neutral perspectives based on their attitude towards given comments. Previously, it was used for lexical or syntactic feature extraction, and each document or text unit was assigned as a polarity label. Opinion mining is used to analyze customer feedback in the form of reviews. This allows to compare various rival brands. Unstructured text data is mined and a structured summary is provided to clients for comparison. Opinion mining is a technology that uses user opinions, which are submitted and expressed as reviews and online reviews on blogs and forums. This might also be a tweet about a product. [10].

II. RELATED WORKS

SVM, Naive Bayes, and backpropagation methods are used to express the emotions from the dialogue. SVM methods outperform Naive Bayes and KNN, and the training dataset produces a lot of positive feedback [11]. This is a type of natural language processing (NLP) model, which is used to track public awareness of a law, policy, or marketing campaign. This includes a legal information and legal reviews and how opinions are collected and reviewed. Posting at social media, creating policies, etc. [12].

Sentiment analysis plays a vital role in the decisionmaking process of services, products, and social problems

Opinion mining includes the concept of information retrieval. Positive and negative reviews are divided into two steps. The product database review is used to find the source dataset for tagging the speech and NLP toolkit.

Before buying a product, people usually learn about themselves by reading more online reviews. To make more profit, sellers often try to forge more user experience. Several ways to analyze product review datasets. We also propose an emotion classification algorithm for supervised learning of product reviews in two different datasets. Our experimental approach looks at the accuracy of all emotion classification algorithms and how to determine which one is more appropriate. This system uses a dictionary-based unsupervised technique.

Lexicon is used to look up thesaurus for words that convey a point of view. The proposed strategy requires mining and interpreting customer feedback [7]. Collect reviews from product users and reviewers and use them as input to the system. There are various mining methods for data extraction. There are various words that are flagged, but some of them indicate that these words appearing in the reviews are not important when performing sentiment analysis [14]. These words that don't make sense in a sentence are called stopwords and should be removed from the comments.

Using text analytics, the above strategy isolates utterances into their constituents of speech [2]. The system's output assists users in making judgments and makes it simple to determine how many positive and negative documents exist. The overwhelming view's phrases are used to estimate a document's polarity.

Each document is provided a positive, negative rating by the system, and the total number of significant and undesirable review documents is indicated in the output. The system's output assists users in making judgments and makes it simple to determine how many good and negative reviews are available. The majority view's terminology is often used to determine a document's polarity.

Artificial neural networks have been used to calculate POS tags successfully and with good results. Part of speech (POS) tagging is a standard language processing technique that combines evaluating meaning of a text (legislature) with a certain word or phrase depending on its definition and context alongside tagging them as that part-of-speech.

III. EXISTING SYSTEM

Online reviews have become increasingly essential in recent years when it comes to purchasing purchases. This is due to the fact that these reviews can give buyers a wealth of information about the product or service. Spammers, on the other hand, may fabricate and publish bogus reviews in order to promote fictitious items or services or to reduce the quality of those products or services [21]. Customers will be misled and make poor decisions as a result of spammers' actions. As a result, spotting fraudulent (spam) reviews is a major challenge. Opinion spamming is the practice of using excessive and illegal means, such as posting a huge number of phony reviews, to generate skewed good or negative opinions about a specific product or service in order to promote or demote it. Fake, spam, or phony reviews are made for this purpose, and the authors responsible for creating such false information are known as fake or spam reviewers [20].

IV. PROPOSED SYSTEM

Various methods for detecting comment spam have been investigated to improve the accuracy and use of opinion mining [15]. It provides a complete description of existing strategies for determining if a comment is spam. The process of classifying words into audio parts. Also known as the audio part or vocabulary category. NLP Techniques are used for monitoring to understand the problems raised in social media. POS is an unsupervised learning method. POS linguistically divides sentences by grammatically labeling categories. Each word in a sentence is marked with extensions such as nouns, adjectives, verbs, and prepositions. Compare the POS-tagged output with negative

and positive datasets. Good if the positive value is exceeded or bad. Each document is given a positive, negative, or neutral rating, and the output shows the total number of successful, negligible documents [22]. In order to acquire more accurate results from Opinion mining, other techniques such as IP Address Tracking and Ontology are used to detect spam reviews [19]. After detecting spam reviews in the original Dataset, a new Dataset without spam reviews is established, and opinion mining is performed on the new Spam Filtered Dataset. Finally, a new technique is proposed for more precisely detecting spam reviews and performing opinion mining using spam-filtered data [17].

V. FEATURE EXTRACTION

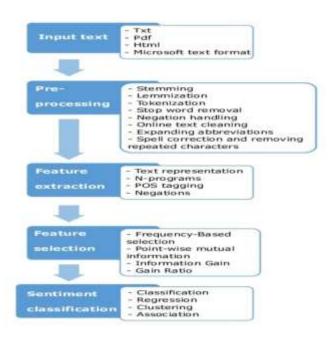


Fig.1.Diagram for feature Extraction

VI. SYSTEM DESIGN

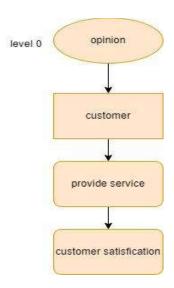


Fig.2.customer service

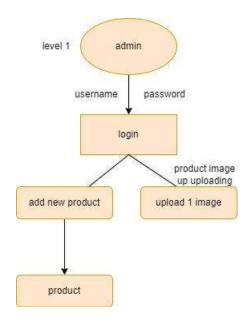


Fig.3.admin service

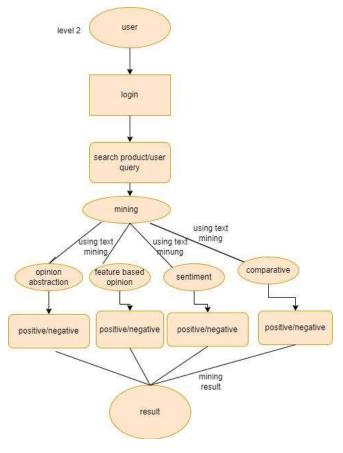


Fig.4.product feedback

VII. MODULES EXPLANATION

First user want to enter mail id and otp number for enter the

system.After accessing the system, users can view the product and also post the review .User id will be verified for posting the reviews when the admin can see the review as

spam they have access to block the email id and also admin will delete the review which is fake[18].

In admin login, they can login using admin Id and Password, and also they will add the product in the system. In user login, the user will login by using user ID and Password and also the users can post reviews about the product.

VIII. METHODS INVOLVED IN NLP

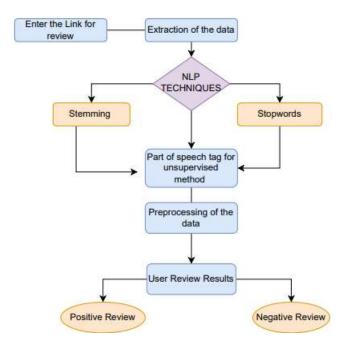


Fig.5.Process of Opinion Mining

A. Data Extraction, Stop Words

Information extraction is the act or practice of collecting data from a data stream for more palletization. Therefore, after importing into an intermediate extraction system, data transformations are typically performed and metadata may be added before exporting to another stage in the data workflow. [3].

This stopword has various flagged words, and there are also reviews of words that are not important when performing sentiment analysis. These words that don't make sense in a sentence are called stopwords and should be removed from the comments. Common words such as "is, a, the" marked "_ IS", "_ BE", "_ CC", "_ TO" to further reduce complexity has been deleted.

With Stop Words	Without Stop Words
growing-up-with-hearing-loss/	/growing-hearing-loss/
coming-to-terms-with-hearing-loss/	/coming-terms-hearing-loss/
the-world-of-being-hearing-impaired/	/world-being-hearing-impaired/
echo-in-ears-from-talking-people/	/echo-ears-from-talking-people/
listening-is-exhausting/	/listening-exhausting/
living with-hearing-loss/	/living-hearing-loss/
/what-is-hearing-loss/	/what-bearing-loss/

Fig.6.Stop words

B. Stemming, POS tagging

By subtracting linguistic features, the separating technique is utilized to retrieve the common structure of the words. Cutting down a tree's branches to its stems is known as stemming. Scratching is a tactic used among web pages to crawl a string[16]. For all of this objective, a browser can only maintain the cuttings. As a response, the index levels are lower and retrieving frequency is enhanced.

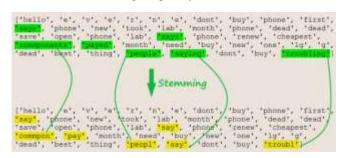


Fig.7.Stemming

NLP should be used in Image captioning to sort phrases into lexical items. POS linguistically marks categories by grammatically splitting sentences. Each word in a sentence is marked with an extension, such as a noun, adjective, verb, or prepositions.

Parts of Speech(POS) Tagging

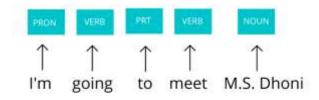


Fig.8.POS Tagging

C. Data Comparison

Using POS tagger, match outcomes to adaptive and maladaptive statistics. It is excellent if the favorable aim is fulfilled. Otherwise, it is bad. So, the output of POS tagging is compared with positive and negative comments.

D. Data Preprocess

1. URL and Hash tags

The preprocessing of the data is displayed if both positive and negative values are the same intermediate value. Utilizing Web Pages and captions, users power produced information on the issue. All URLs and hashtags are removed from comments under the proposed system.

2. Lower case, Identifying Punctuation

User comments may contain uppercase and lowercase letters. Uneven use can also give meaning to a word. To reduce ambiguity, the proposed system further processes the data by converting all reviews to lowercase [8]. Punctuation and whitespace are identified and removed to avoid redundant features and other conflicts.

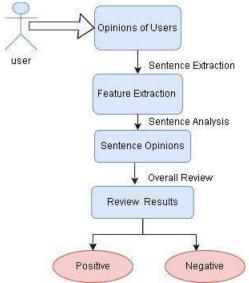


Fig.9. Steps Used in the Process

IX. RESULT

Review sites, forums, and blogs are rich in user-created content, making opinion mining an attractive area of research. Opinion mining is used in a variety of settings, including market research, decision-making, advertising. Companies can measure product acceptance and establish plans to improve their products through opinion mining. Individuals can also use opinion mining tools to make purchase decisions based not only on specifications, but also on user experience and public opinion. In this paper, using unsupervised techniques of POS tagging we show positive and negative comments given by customers in eCommerce sites. The negative words are identified in the review using a negative dictionary. The reviews are added in spam, if it has more than five negative words.



Fig.10.Book Store Login



Fig.11.Home Page For User



Fig.12.User Purchasing Product



Fig.13.Product for sales



Fig.14.output screen



Fig.15. Diagram for Stopwords



Fig. 16. Preprocess of Product Review

X. CONCLUSION AND FUTURE WORK

This analysis study introduces Opinion Mine [4], Associate in Nursing IP methodology for opinion mining on problems associated with client reviews. Nowadays, folks area unit victimization on-line feedback from their customers and most web-based and mobile-based applications. Automatic analysis of those on-line reviews can facilitate improving the standard of service. The complete on-line review of consumers on e-commerce sites is evaluated during this study.

In this implementation work, we have a tendency to plan an answer for analyzing the reviews mechanically by applying

unattended machine learning approaches [6].

unlabeled knowledge trumps typical reinforcement ways once it involves decision making assertions in uncommon ways. The vital feature is that it may be custom-made terribly simply to several reviews in ecommerce sites. Our approach may be additionally used for the other social networks like

Amazon, Flipkart, Snapdeal and every one alternative security websites [9]. What is more, we have a tendency to area unit victimization of some product datasets to search out precise reviews given by the purchasers.

REFERENCES

- $\label{eq:chen} \begin{tabular}{l} [1] W. Chen, X. Li, W. Zuo \ and M. Yin, "A survey of sentiment analysis in social media", data and data Systems, pp. 1-47, 2018. \end{tabular}$
- [2] S. Sun, C. Luo and J. Chen, "A review of language process techniques for opinion mining systems", Inf. Fusion, vol. 36, pp. 10-25, 2017
- [3] E. Sarr, S. Ousmane and A. Diallo, "FactExtract: Automatic assortment and Aggregation of Articles and print media Factual Claims from on-line Newspaper", 2018 Fifth International Conference on Social Networks Analysis Management and Security (SNAMS), pp. 336-341, 2018.
- [4] A. Ligthart C. Catal and B. Tekinerdogan "Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification" Appl. Soft Comput. vol. 101 Mar. 2021.
- [5] B. B. Ahamed and K. Murugan, "Study of Socio-Linguistics on-line Review System victimization Sentiment grading Method", International Conference on Intelligent Computing & optimization, pp. 569-580, 2019, October.
- [6] R. Barbado O. Araque and C. A. Iglesias "A framework for fake review detection in online consumer electronics retailers" Inf. Process. Manage. vol. 56 no. 4 pp. 1234-1244 Jul. 2019.
- [7] M.S.Kumar, Dr. A. Bala, "Analyzing Twitter Sentiments Through massive Data", 2016 International Conference on Computing for property world Development (INDIA Com).
- [8] P. Kalaivani, Dr. K.L. Shunmuganathan, "Sentiment classification of showreview by supervised machine learning approach", Indian Journal of applied science and Engineering (IJCSE) Vol. 4 No.4 Aug-Sep 2013.
- [9] F. Gräßer S. Kallumadi H. Malberg and S. Zaunseder "Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning" Proc. Int. Conf. Digit. Health pp. 121-125 Apr. 2018.
- [10] S. Kataria and M.D. Singh, "A Review of information Classification victimization K Nearest Neighbour rule," International Journal of rising Technology and Advanced Engineering, vol. 3, no. 6, pp. 354-360, June 2013.

- [11] Prameswari, P., Surjandari, J., & Laoh, E. (2017). Opinion mining from on-line reviews in Bali traveler space. In the third international conference on science in info technology.
- [12] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, T. By, "Sentiment Analysis on Social Media", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 919-926, 2012
- [13] Ravi, K., & Ravi, V. (2017). A survey of opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems, (89), 14–46.
- [14] A. A. Farisi Y. Sibaroni and S. Al Faraby "Sentiment analysis on building reviews victimization multinomial naïve mathematician classifier" Journal of Physics: Conference Series vol. 1192 no. 1 pp. 012024 2019. [15] R. Geetha, P. Rekha and S. Karthika, "T witter Opinion Mining and Boosting victimization Sentiment Analysis," 2018 International Conference on pc, Communication, and Signal process (ICCCSP), 2018, pp. 1-4, doi: 10.1109/ICCCSP.2018.8452838.
- [16] M. Wongkar and A. Anglesey, "Sentiment Analysis victimization Naive mathematician rule Of the info Crawler: Twitter," 2019 FourthInternational Conference on information processing and Computing (ICIC), 2019.
- [17] H.Aghakhani A. Machiry S. Nilizadeh C. Kruegel and G. Vigna "Detecting deceptive reviews using generative adversarial networks" Proc. IEEE Secur. Privacy Workshops (SPW) pp. 89-95 May 2018.
- [18] Z. You T. Qian and B. Liu "An attribute enhanced domain adaptive model for cold-start spam review detection" Proc. 27th Int. Conf. Comput. Linguistics pp. 1884-1895 2018.
- [19] F. Khurshid Y. Zhu Z. Xu M. Ahmad and M. Ahmad "Enactment of ensemble learning for review spam detection on selected features" Int. J. Comput. Intell. Syst. vol. 12 no. 1 pp. 387-394 2018.
- [20] N. A. Patel and R. Patel "A survey on fake review detection using machine learning techniques" Proc. 4th Int. Conf. Comput. Commun. Autom. (ICCCA) pp. 1-6 Dec. 2018.
- [21] X. Tang T. Qian and Z. You "Generating behavior features for cold-start spam review detection with adversarial learning" Inf. Sci. vol. 526 pp. 274-288 Jul. F. Xia J. Liu H. Nie Y. Fu L. Wan and X. Kong "Random walks: A review of algorithms and applications" IEEE Trans. Emerg. Topics Comput. Intell. vol. 4 no. 2 pp. 95-107 Apr. 2020.2020. [22] X. Wang W. Jiang and Z. Luo "Combination of convolutional and
- [22] X. Wang W. Jiang and Z. Luo "Combination of convolutional and recurrent neural network for sentiment analysis of short texts" Proc. 26th Int. Conf. Comput. Linguistics: Tech. Papers (COLING) pp. 2428-2437 2016
- [23] J. Li M. Ott and C. Cardie "Identifying manipulated offerings on review portals" Proc. Conf. Empirical Methods Natural Lang. Process. pp. 1933-1942 2013.