Sentimental Analysis and Opinion Mining on Online Customer Review

Anushree Raj[i], Joylin Denita Dsouza[2,] Shabarish S.K[3]

[1] *Assistant Professor- IT Department, AIMIT, Mangaluru, anushreeraj@staloysius.ac.in*
[2] *MCA Student, AIMIT, Mangaluru, 2117026joylin@staloysius.ac.in*
[3] *MCA Student, AIMIT, Mangaluru, 2117103shabarish@staloysius.ac.in*

---***---

Abstract: *Customer Opinions play a very crucial role in daily life. The views of other people are taken into account when we must make a decision. Today, a large number of internet users share their ideas about a variety of products on blogs, review websites, and social networking sites. The number of individuals buying things from the web is growing, and there are an ever-increasing number of results kept there. As a result, the number of user evaluations or postings is growing daily. Business organizations and corporate organizations are constantly interested in hearing what consumers or individuals have to say about their goods, services, and support. Today, if someone needs to purchase an online item, he/she can first view other buyer's reviews on the product site and take the right decision accordingly. There are many opinions on the web about a product so taking decisions might be difficult. Therefore, opinion mining is used for classifying the reviews according to their polarity. The method of extracting opinions from reviews is known as opinion mining (OM). Any user who wants to choose a product or business should consider consumer reviews when doing so. Sentiment analysis is another name for opinion mining. Our main goal is to develop a system for analyzing opinions, which implies judging various consumer products.*

Key words: *Opinion Mining, Sentiment Classification, Online Reviews, Data Mining, Web Mining,* *Sentiments analysis (SA).*

1. INTRODUCTION

The current era is characterized by the presence of Ecommerce retailers everywhere around us. Almost all company phases are E-commerce stores. With widespread Internet connection and knowledge of the method, the Ecommerce business has soared to new heights in recent years. There are several criteria that contribute to an Ecommerce store's success and credibility. Product Reviews, on the other hand, are a significant component in raising the reputation, standard, and assessment of an Ecommerce company [1]. Product Reviews are one of the most significant tools accessible to an Ecommerce shop, namely Customer Feedback.

One critical duty for the Ecommerce company is to keep up its reputation in the web industry. Naturally, it takes a lot of effort to get that reputation, but it costs very little to lose it: Product Reviews are the most ideal approaches to keep up their series of wins. Item reviews and criticisms have modified the enjoyment for the internet industry because the web has become so widespread. Product Reviews are the elements that determine the customer's trustworthy relationship with the business - they assist build dependability and trust and explain the possible buyer the item significantly more clearly and the viewpoints that distinguish it from whatever is left of the things anywhere else.

An Ecommerce shop with a good collection of client reviews for its products displays a good reputation among customers. Presently the opinions of others regarding a product influence purchasing decision [2]. For example, the consumer will just purchase the item by reading the feedback written by customers. he obtains clear thinking about the determination and effectiveness of the organization's product specifics and intricacies to their products However, in real-life situations, half of The highlights that the producer gives about the item are not true. As a result, only authentic customers who use that product may provide specific information about the product Now arrives the importance of reviews We are currently witnessing financial waste as a result of poor purchase decisions.

The presentation of semantic analysis on reviews tackles the above problem. Users' premium is steady just in brief period. so, client subjects from surveys can be delegate for e.g. if there should arise an occurrence of a versatile phone, different individuals have distinctive ideas. a few people focus on camera, where some focus on battery reinforcement thus on. They all have customized territory of enthusiasm for the item. The importance of sentiment analysis comes here. Sentiment analysis otherwise known as opinion mining is the process of determining the emotional tone behind a series of words. Sentiment analysis is extremely useful in online e-commerce sites to monitor the reviews it allows us to gain an opinion about the product. Using sentiment analysis on product reviews helps us to extract the emotional tone towards the product. Through natural language processing and machine learning. product reviews in e-commerce sites are written in natural languages such as English. This technique is used to figure out the sentiment or emotion associated with the underlying text. So, if you have a piece of text and you want to understand what kind of emotion it conveys, for example, anger, love, hate, positive, negative, and so on you can use the technique sentimental analysis

2. OBJECTIVES

The purpose of sentiment analysis and opinion mining is to teach computers to understand and communicate emotion. This work focuses on mining reviews from websites such as Amazon, which allow users to freely publish their opinions. The reviews are automatically extracted from the website. Most semi-structured online data contain a wealth of helpful information. With the increasing growth of ecommerce, more and more things are being sold online. People are increasingly doing their purchasing on the internet. Not only that, but consumers frequently discuss their product-related experiences on the internet. They utilize blogs, Twitter, and other such websites to express themselves and share their experiences with others. Websites for social interactions have grown in popularity. It has become normal practice for online retailers to allow their customers to evaluate or voice thoughts about the things that they have purchased in order to improve customer happiness and shopping experience.

As the number of online buyers grows, so does the number of reviews express on the internet. Hundreds of thousands of reviews have been written about some products. Understanding the consumers' perceptions of products is extremely beneficial to both merchants and customers who want to purchase those products in the future. When the quantity of reviews is huge, reading them all one by one is inefficient. The review material also sometimes generates confusions. The majority of product reviews have a lot of long sentences. Few of them truly express their opinion. As a result, it is more difficult to read and comprehend comments.

If someone reads only a few reviews and then makes a judgement, the decision may be prejudiced. Because of these factors, having a better data mining technique to mine these semi-structured product reviews is critical. Not only are product reviews significant, but so are reviews on certain places, sports, and movies if they are mined correctly to get their real viewpoint. Many researchers have looked into this issue in recent years [3]. The field of study is known as opinion mining and sentiment analysis. This research area has two primary goals. They are (1) locating product features on which reviewers have remarked and (2) determining whether the remarks are positive or negative. Both activities are extremely difficult, and several studies have been undertaken on them.

3. LITERATURE REVIEW

Rajkumar et al. rendered ML approaches say Naïve Bayes (NB) and SVM for performing SA on reviews of a specific product. In those approaches, the dataset was gathered as of Amazon, which comprised reviews regarding Laptops, Cameras, Mobiles, Tablets, video surveillance, and TVs. Subsequently, stemming, stop word removal, and also punctuation marks removal was executed and it was transmuted into a bag of words. This dataset was contrasted to opinion lexicons, that is, 4783 negative and 2006 positive words with sentiment scores intended for every sentence were evaluated. Utilizing score and disparate features, the NB along with SVM were employed and diverse accurateness was computed. The ML approaches proffered the good outcomes to categorize product reviews. NB got 98.170% accuracy and SVM got 93.54% accuracy for Camera—related Reviews. The approach utilizes the SVM, which encompasses several key parameters that are required to be set properly for attaining the best classification outcomes. Thus, the SVM renders lower accuracy in classification [4].

Satuluri Vanaja and Meena Belwal rendered an Aspect-level SA, which was attained by Identification, aggregation, and Classification. The pre-processing includes Parts-of-Speech tagging to every word in each sentence, extracting frequently used words, removing stopping or unwanted words and adjective extraction from the sentences. The classification was executed utilizing ML algorithms, NB and SVM classification algorithm and the performance were contrasted centered on Recall, F1 measure, and Precision. The outcomes evinced that it attained more accurateness from the NB when weighed against the SVM. The SVM approach was not apt for large datasets [5].

Wei Zhang et al. propounded an emotion classification algorithm grounded on SVM as well as latent-SA (LSA). Primarily, Psychology and NLP were integrated to divide the emotions in the online reviews onto '4' categories: a. happiness, b. hope, c. disgust, and d. anxiety. Subsequently, the LSA approach was utilized for optimizing the text feature extraction and employed the SVM as a classifier for ameliorating the emotional classification accurateness and computational efficacy. The experiential outcomes evinced that the model could effectually compute online reviews. Context meanings of data with DL algorithms were utilized for combining the reviews' theme, sentiment classification and product characteristics for further enhancing the multiple class emotional detection accuracy. The approach employs only a less amount of data for analyzing, which is not efficient [6].

Barkha Bansal and Sangeet Srivastava rendered a Hybridized Attribute-centric Sentiment Classification (HABSC) for infusing domain-specific knowledge and collecting the implicit word relations. This approach found the utmost frequent bi- gram as well as tri-gram in the corpus, followed by POS tagging for retaining opinion words and aspect descriptions. Subsequently, it has deployed TFIDF for signifying every document, followed by automatic extraction of an optimal topic. All the adverbs and adjectives were labelled utilizing pre-existing lexicon and domain-related knowledge. The method's efficacy was tested utilizing datasets. The outcomes evinced that the classification accurateness of HABSC exceeded for disparate methods and also evinced less computational time as contrasted to distributed vectorization frameworks. The approach was not

effective in detecting the attribution, and it has a high computational time [7].

Chonghui Guo et al. examined a ranking approach via online reviews grounded on diverse aspects of variant products that integrated the subjective as well as objective sentiment values. Primarily, the product's sentiment value was evaluated by ascertaining the weights of those aspects with the LDA topic design. At the time of this process, the realistic meaning of every single aspect was as well summarized. Subsequently, consumers' personalized preferences were regarded whilst evaluating the total scores of variant products. Meanwhile, comparative superiority in-between every '2' products also added into final scores. By utilizing the Page Rank algorithm, the attained final score of every product was evaluated as of the constructed graph. The outcome elucidated that whilst regarding only objective sentiment values of the product, the ranking outcome attained by this approach had a good correlation with the primary sales orders. But the system used the LDA, which was sensitive to overfit, and validation of LDA models was at least problematic [8].

Sumbal Riaz et al. recommended an approach termed text mining for examining customer reviews to ascertain the customers' opinions and executed the SA on the massive dataset of product (6 sorts) reviews proffered by disparate customers on the internet. In this approach, SA was employed at the phrase level instead of document-level for computing every term's SP. Then key graph keyword extraction was used aimed at extracting keywords as of each document with high-frequency terms and the intensity of SP by gauging its strength was evaluated. The k-means clustering was utilized for grouping data on the base of sentiment strength value. Those values were contrasted to the star rating of the same data and the excellent and neutral sentiment toward products was examined. The approach uses clustering which may bring about over clustering [9].

4. METHODOLOGY

This research paperwork is divided into modules. This research study focuses on Amazon product review mining that follows the free review structure. There will be no constraints on the user's ability to write a review. Users of the online shopping site Amazon are encouraged to leave product reviews on the items they purchase. Amazon uses a 1-to-5 scale for all products, regardless of category, making it difficult to assess the benefits and drawbacks of various elements of a product. It explains how to extract product attributes from opinion statements. It employs a SentiWordNet-based algorithm to determine the sentence's opinion.

Sentiment analysis

Sentiment analysis is the automated method of deducing a person's feelings on a specific topic from written or spoken language. Sentiment analysis is sometimes known as opinion mining, and it is a branch of natural language processing that extracts hidden opinions in text.

In extracting an expression, there are three attributes to consider:

a) polarity- what kind of polarity the customer expresses in his review; it can be positive, negative, or neutral.

b) subject- the thing being discussed.

c) opinion holder- the customer who expresses an opinion about a product through reviews.

Because of its multiple functional uses, sentiment analysis is currently a subject of extraordinary priority and advancement. Because the amount of freely and covertly accessible data on the Internet is always rising, a large number of works describing feelings are available through review sites, conversations, web journals, and social media. With the assistance of hypothesis examination frameworks, this unstructured data could be transformed into ordered knowledge on public sentiments regarding things, administrations, brands, legislative difficulties, or any subject about which people can express opinions [10]. This data can be used for commercial applications such as showcasing analysis, advertising, item surveys, net advertiser score, item input, and client management.

Scope of Sentiment Analysis

Sentiment analysis can be used at various levels of detail:

- **Document level** sentiment analysis obtains the sentiment of a complete document or paragraph.

- **Sentence level** sentiment analysis obtains the sentiment of a single sentence.

- **Sub-sentence level** sentiment analysis obtains the sentiment of sub-expressions within a sentence.

The kind of sentiment analysis

There are different types of sentiment analysis where in this system we propose a combination of fine-grained sentiment analysis, emotion detection, and aspect-based sentiment analysis.

Fine-grained Sentiment Analysis

Instead of only looking at generic viewpoints, we are now delving further into opinion mining. Instead of taking positive, neutral and negative opinions can considers the following categories: Very positive

- Positive

- Neutral

- Negative

- Very negative

We can also utilize star representation, with 5 stars representing a highly good view and 1 star representing a very negative one.

Emotion detection

Emotion detection aims at detecting emotions like, happiness, frustration, anger, sadness etc. in the reviews. Just like mining the opinion from the review emotions also has its importance to form precise sentiment about a product.

Aspect-based Sentiment Analysis

In this type of sentiment analysis, we discuss not only the sentiment of the review, but also which specific aspect or feature of the product we have an opinion about. For example, "the battery life of the mobile phone is insufficient." The sentence expresses a bad view about the mobile phone, more specifically about the battery life, which is a feature of the phone.

Working of sentiment analysis

There are several approaches and algorithms to implements sentiment analysis systems, which can be classed as:

- **Rule-based** systems that execute sentiment analysis based on a set of manually written rules.

- **Automatic systems** that learn from data using machine learning techniques.

- **Hybrid systems** that combine both rule based and automatic approaches.

Opinion

The information in the text can be divided into two categories: facts and opinions. Whereas facts are objective statements, opinions are subjective expressions that include user attitudes and feelings about the product [11].

Like other NLP challenges, sentiment analysis can be classified as a classification problem with two subproblems to solve-They are:

Subjectivity classification-classifying the sentence into subjective or objective

Polarity classification- classifying the sentence opinion into positive, neutral and negative

In an opinion, the element the content discussions about can be an item, its segments, its aspects, its characteristics, or its

highlights. It could likewise be an item, an administration, an individual, an association, an occasion, or a subject. As an example, take a look at the opinion below:

"The battery life of this mobile phone is excessively short. "A negative feeling is communicated about an element (battery life) of a substance (mobile phone).

Direct vs. Comparative Opinions

There are two sorts of opinions: direct and comparative. Direct conclusions give a sentiment about a substance straightforwardly, for instance: "The sound quality of mobile phone A is poor." This direct opinion states a negative sentiment about mobile phone A. In comparative feelings, the opinion is communicated by contrasting a substance and another, for instance: "The sound quality of mobile A is better than that of mobile B."

Among the different approaches to sentiment analysis accessible, (SA), only two major categories are prevalent. The first category SA's problems are solved by implementing the machine

method of learning Several techniques are used in this group. used in an attempt to extract significant traits that more provide precise information regarding the polarity of sentiments

As the procedure progresses, the technique employed is regularly monitored. A carefully annotated corpus is required. The second group employs a method that is more linguistically inclined.

referred to as the lexicon-based approach according to the source. The investigation begins with words or sentences. exhibiting semantic polarity qualities.

There are some a variety of machine learning (ML) methodologies They have created for the goal of categorizing literature as positive or bad or classes that are positive. The approaches' performances: Naive Bayes (NB), Support Vector Machines (SVM), and Maximum Entropy (ME) and classification are used. really successful ID3 is one of the other ways. Centroid Classifier,

Winnow Classifier, and K-Nearest Neighbor Mining Approach for Neighbor and Association Rules.

Naïve Bayes (NB) classification method is commonly utilized for classifying text documents.

This technique is based on a probabilistic model and employs cooperative probabilities of certain terms and their respective group for the estimation of the probability of a certain group, with a text document as input.

In addition, the Support Vector Machine (SVM) is being presented as a classifier to handle the challenges of pattern

recognition between two groups. The Support Vector Machine (SVM) seeks to determine the optimal hyperactive plane margin separation between two groups of data. It was designed to solve separable instances, but it can be expanded to handle linearly non-separable problems by transferring the original data vector to higher-dimensional spaces. Many researchers believe that the SVM classifier is the best method to utilise for text classification and have used it.

5. ANALYSIS STUDY:

Below are some of the primary uses for sentiment analysis and opinion mining.

1) Identifying opinion spam: Customers may publish evaluations regarding products with malicious intent. These reviews may be divided into "not spam" and "spam" material using sentiment analysis and opinion mining.

2) Buying a Product or Service: Using this method, consumers may quickly assess other people's opinions and experiences with any Product or Service and compare rival brands.

3) Product or service quality improvement: Manufacturers may use this to gather both positive and negative feedback about their goods, which helps them raise the standard of their goods.

4) Marketing analysis: With their new government strategy, products or services be examined. These are all outcomes that may be attributed to group intelligent research.

5) Policy Making: Through sentiment analysis, policy makers may learn how citizens feel about a certain policy and use this knowledge to develop new, better policies that are more geared toward the needs of citizens.

6) Decision-Making: Individuals' perspectives and backgrounds are a crucial component in the decision-making process. It offers assessed people's opinion that may be properly employed for making decisions

6. CONCLUSION

In conclusion, using sentiment analysis or opinion mining to mine a variety of unstructured data has emerged as a crucial research issue. Sentiment analysis helps to create better goods, services, and effective business management. In the review article, related work in the field of sentiment analysis from the period was given.

Future study is required to further improve the performance measurements. For any new applications that follow the principles of data mining, sentiment analysis or opinion mining can be used.

Even while the algorithms and techniques used for sentiment analysis are improving quickly and producing high-quality findings, many issues in this area of study are still open, and it might be challenging to spot false reviews just by reading them. Sometimes fake reviews are mistaken for genuine ones and are altered so that no one can tell what their true intentions were. Therefore, the identification of false reviews is another crucial area that calls for deep data mining approaches.

REFERENCES:

[1] E. A. Stepanov and G. Riccardi, "Detecting General Opinions from Customer Surveys," 2011 IEEE 11th International Conference on Data Mining Workshops, 2011, pp. 115-122, doi: 10.1109/ICDMW.2011.63.

[2] Pankaj, P. Pandey, Muskan and N. Soni, "Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 320-322, doi: 10.1109/COMITCon.2019.8862258.

[3] P. Kherwa, A. Sachdeva, D. Mahajan, N. Pande and P. K. Singh, "An approach towards comprehensive sentimental data analysis and opinion mining," 2014 IEEE International Advance Computing Conference (IACC), 2014, pp. 606-612, doi: 10.1109/IadCC.2014.6779394.

[4] Rajkumar S Jagdale, Vishal S. Shirsath, Sachin Deshmukh, "Sentiment Analysis on Product Reviews Using Machine Learning Techniques: Proceeding of CISC 2017", Cognitive Informatics and Soft Computing, Advances in Intelligent Systems and Computing 768,https://doi.org/10.1007/978-981-13-0617-4_61.

[5] Satuluri Vanaja, Meena Belwal, Aspect-Level Sentiment Analysis on E-CommerceData,2018.

[6] Z. Zhang, H. Li and W. Yu, "Fine-grained opinion mining: An application of online review analysis in the express industry," 2017 3rd IEEE International Conference on Computer and Communications (ICCC), 2017, pp. 1498-1503, doi: 10.1109/CompComm.2017.8322790.

[7] Barkha Bansal, Sangeet Srivastava, "Sentiment classification of online consumer reviews using word vector representations", Procedia Computer Science, Volume 132, 2018.

[8] Chonghui Guo, Zhonglian Du, Xinyue Kou, "Mining Online Customer Reviews for Products Aspect-Based Ranking", International Symposium on Knowledge and Systems Sciences, October 2017,DOI: 10.1007/978-981-10-6989-5_13.

[9] Sasikala p, Mary Immaculate, Sheela Lourdusamy, "Sentimental Visualization: Semantic Analysis of Online Product Reviews Using Python and Tableau," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 1-3, doi: 10.1109/BigData50022.2020.9391769.

[10] C. S. R. Priya and P. Deepalakshmi, "Study on Online Review Based Consumer Sentimental Analysis using Machine Learning Approaches," 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), 2022, pp. 610-616, doi: 10.1109/AIC55036.2022.9848932.

[11] P. K. Singh, A. Sachdeva, D. Mahajan, N. Pande and A. Sharma, "An approach towards feature specific opinion mining and sentimental analysis across e-commerce websites," 2014 5th International Conference - Confluence the Next Generation Information Technology Summit (Confluence), 2014, pp. 329-335, doi: 10.1109/CONFLUENCE.2014.6949312.