iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** X **Month of publication:** October 2023

DOI: https://doi.org/10.22214/ijraset.2023.56255

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue X Oct 2023- Available at www.ijraset.com

A Survey on "Sentiment Analysis of E-commerce Website's Reviews"

Miss. Yashashri Gayki[1], Mr. Pratham Chatke[2], Miss. Rani Nandane[3], Mr. Umesh Shinde[4], Mr. Manav Ghugare[5]
Priti Rathod[6], Prof. Minal Pardey[7]
[1, 2, 3, 4, 5, 6]*Student, Department of Computer Engineering, JCOET, Yavatmal*
[7]*Asst Prof. Department of Computer Engineering, JCOET, Yavatmal*

Abstract: ***Sentiment analysis, a vital component of natural language processing, has gained significant relevance in the realm of ecommerce websites. In this digital age, where consumers heavily rely on online reviews to inform their purchase decisions, understanding and harnessing sentiment in ecommerce reviews is paramount. This abstract explores the utilization of sentiment analysis techniques to extract valuable insights from customer feedback, offering a panoramic view of its applications, challenges, and implications. We delve into keyword extraction, sentiment polarity classification, and the integration of sentiment analysis into recommendation systems. This paper also examines the evolving role of sentiment analysis in enhancing user experiences, brand reputation management, and product development. By decoding the sentiments hidden within ecommerce website reviews, businesses can strategically adapt, improve customer satisfaction, and thrive in a highly competitive online marketplace.***

Keyword: Sentiment analysis, Python, Machine Learning, Natural Language Processing (NLP), Text analysis

I. INTRODUCTION

Sentiment analysis involves using natural language processing and computational linguistics to extract subjective information and determine the attitude or emotional content in a piece of writing. This process categorizes the sentiment as positive, negative, or neutral. However, relying on humans for this task is impractical due to its time-consuming and costly nature, especially when dealing with a large volume of customer reviews. Semantria's cloud-based sentiment analysis software breaks down a document into its parts of speech (POS tags) and identifies sentiment-bearing phrases using a specially designed algorithm. Each sentiment-bearing phrase is then scored on a logarithmic scale from -10 to 10, and these scores are combined to determine the overall sentiment of the document or sentence, ranging from -2 to 2. Existing approaches to sentiment analysis include keyword spotting, lexical affinity, statistical methods, machine learning, and natural language processing. Sentiment analysis, also known as opinion mining, is a prominent area of research in natural language processing, focusing on analyzing people's opinions, attitudes, and emotions expressed in written language. This analysis can be achieved using either machine learning or a lexicon-based approach. In this study, a lexicon-based approach was applied, which is a feasible and practical method capable of analyzing review text without requiring machine learning training. However, it's important to note that statistical text classifiers may struggle to achieve high accuracy with smaller text units like sentences or clauses, performing more effectively at the page or paragraph level.

II. LITERATURE REVIEW

A method was developed for ranking online product reviews based on different aspects of the products[01]. This method takes into account both objective and subjective sentiment values. They used a technique called Linear Discriminate Analysis (LDA) to determine how important various aspects of the product are, which helps calculate the objective sentiment value. They also considered individual consumer preferences to calculate overall scores for different products.

Focusing on three specific features of smart phones: Camera, Battery, Screen, Sounds, Design, and hardware/software performance. They collected reviews related to these features and identified common positive and negative words used in these reviews[02]. They used MATLAB to create a system that counts how often these positive and negative words appear in the reviews to figure out the overall sentiment of the products.

The research highlighted the significance of emoticons in sentiment analysis[03]. The paper briefly discussed various factors affecting sentiment analysis, such as detecting sarcasm, dealing with multiple languages, acronyms, slang, lexical variations, and keeping dictionaries up-to-date. Essentially, the focus was on understanding how emoticons and these factors impact sentiment analysis.

All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

A comprehensive overview of sentiment analysis using Natural Language Processing (NLP) techniques was provided[04]. Sentiment analysis is about determining the sentiment (positive, negative, or neutral) expressed in text. They introduced a new method called "Hybrid Weighted Word2Vec (HWW2V)" for representing text. This approach combines Bag of Words (BOW), Weighted Word2Vec, and sentiment lexicon-based representations to achieve better results in sentiment analysis.

The process of performing sentiment analysis was explained[10]. The focus was on describing the various methods used for sentiment analysis and the tools used to carry out the analysis. Essentially, this paper outlined the steps and techniques involved in understanding and categorizing sentiment in text.

A sentiment dictionary tailored to a specific domain using external textual data can be created[05]. Sentiment analysis often involves categorizing documents based on their opinions. Different classification models have their strengths and weaknesses. Kai Yang proposed a highly effective hybrid model that combines various single models to overcome their individual weaknesses. Their experiments showed that this hybrid model performs better than using single models alone.

A clever way was figured out to determine how people feel about things[06]. They used a mix of computer smarts and a tool that knows about feelings. They tested their method using messages from Twitter, where people talk about all sorts of stuff. They also looked at a special list of words to tell if a message was happy, sad, or just plain neutral. To make sure their method worked well, they checked it with some other computer tricks too.

An experiment was done to see which computer method works best for understanding feelings in text[07]. They tried out some traditional computer tricks and some newer ones that work a bit like how our brains learn. They collected a bunch of text, cleaned it up, and then used different computer programs to figure out the feelings in it. They compared these programs to see which one did the best job.

A cool idea came to see how people felt about products on Amazon by reading what others wrote in their reviews[08]. They collected lots of these reviews and made them neat and tidy for the computer to understand. They looked for words and ideas that show if someone likes or dislikes something. Then, they checked how good their computer method was by using numbers to see how well it worked.

They looked at many research papers to see how people use computers to understand feelings in tweets[09]. They had a plan for how to do it: first, collect the tweets, then clean them up, find important stuff in them, and finally, teach the computer to understand. They used two sets of tweets - one for teaching the computer and one for testing it. They did all of this using a computer program.

III. PROPOSED APPROACH:

A. *Text Preprocessing*

Text preprocessing in sentiment analysis involves preparing raw text data for analysis by applying various techniques to clean, structure, and transform the text. This process helps improve the accuracy and effectiveness of sentiment analysis algorithms. Common preprocessing steps include:

1) *Data Cleaning:* This involves removing any noise from the text data, such as special characters, HTML tags, or irrelevant symbols.
2) *Tokenization:* Breaking text into tokens, which can be words, phrases, or even subword units like subword tokenization in models like BERT
3) *Lowercasing:* Standardizing text to lowercase to ensure uniformity and avoid treating the same word differently due to case differences.

By applying these preprocessing steps, the text data is transformed into a cleaner and more structured format, ready for further analysis to determine sentiment and opinions.

B. *Text Representation*

Text representation in sentiment analysis involves converting raw textual data into a numerical or structured format that can be used for machine learning models to analyze and classify sentiments. The primary goal is to transform the unstructured text into a structured format that retains essential information for sentiment analysis while being compatible with algorithms that require numerical input. There are several common techniques for text representation:

1) *TF-IDF:* Term Frequency-Inverse Document Frequency assigns numerical values to words based on their frequency in a document relative to their importance in the entire corpus.
2) *Word Embeddings:* Techniques like Word2Vec or GloVe create dense vector representations of words, capturing semantic relationships between them.

Text representation is a crucial step in sentiment analysis, as it transforms text into a format that allows machine learning models to learn patterns and relationships between words and sentiments. The choice of representation technique depends on the dataset, the complexity of the task, and the performance requirements of the sentiment analysis application.

C. Feature Extraction

Feature extraction in sentiment analysis involves identifying and extracting relevant characteristics or attributes from the preprocessed text data. These extracted features serve as inputs to machine learning models, helping them learn patterns and make predictions regarding sentiment (positive, negative, or neutral). Effective feature extraction is crucial for accurate sentiment analysis. Here's a detailed explanation of feature extraction in sentiment analysis:

1) *N-grams:* Capture sequences of words (bi-grams, tri-grams, etc.) to capture contextual information.
2) *Sentiment Lexicon Scores:* Utilize sentiment lexicons like the AFINN lexicon to assign sentiment scores to words in the text.
3) *Domain-Specific Features:* Incorporate features specific to e-commerce, such as product categories, review length, or reviewer credibility.

Feature extraction aims to condense the information in the text into a structured format that machine learning models can effectively utilize. The choice of features depends on the specific task, dataset, and the desired performance of the sentiment analysis system. Experimentation and understanding the domain are essential for selecting the most effective features.

D. Model Building

Model building in sentiment analysis involves constructing and training a predictive system that can accurately classify or analyze the sentiment conveyed by a piece of text. This process requires selecting an appropriate machine learning or deep learning model, preparing the data, training the model, evaluating its performance, and making necessary adjustments for optimal results. Here's a detailed explanation of the steps involved in model building for sentiment analysis:

1) *Machine Learning Models:* Besides Naive Bayes and Support Vector Machines, you can explore Random Forests, Gradient Boosting, or logistic regression for sentiment classification.
2) *Deep Learning Models:* RNNs, Convolutional Neural Networks (CNNs), or Transformer-based models like BERT and GPT can be used for more advanced sentiment analysis.

Model building is an iterative process that may involve multiple rounds of experimentation, fine-tuning, and evaluation to develop an effective sentiment analysis system. The goal is to create a model that accurately predicts sentiment, making it a valuable tool for understanding and analyzing textual data in various domains.

E. Evaluation

Evaluation in sentiment analysis involves assessing the performance and effectiveness of a sentiment analysis model or system. It helps determine how well the model can predict sentiments from textual data and how reliable its predictions are. Effective evaluation is crucial to understand the model's strengths and weaknesses and to make necessary improvements. Here's a detailed explanation of evaluation in sentiment analysis:

1) *Accuracy:* Measures the overall correctness of sentiment predictions.
2) *Precision:* Indicates the proportion of correctly predicted positive sentiments among all predicted positives.
3) *Recall:* Shows the proportion of correctly predicted positive sentiments among all actual positives.
4) *F1-score:* Combines precision and recall into a single metric, helpful for imbalanced datasets.
5) *AUC-ROC:* Useful when dealing with binary sentiment classification and assessing the model's ability to distinguish between positive and negative sentiments.

Evaluation in sentiment analysis provides insights into how well the model performs in real-world scenarios and guides further improvements to enhance accuracy and applicability. It is an integral part of the model development lifecycle, ensuring that the sentiment analysis system meets the desired performance criteria

F. Post Processing

Text post-processing in sentiment analysis refers to the final steps taken after the sentiment prediction has been made by a model. It involves modifying, refining, or enhancing the output to improve its readability, coherence, or usability for the end user. The goal is to present the sentiment analysis results in a format that is more understandable and actionable.

Here's a detailed explanation of text post-processing in sentiment analysis:

1) *Threshold Tuning:* Adjust the classification threshold to optimize the trade-off between precision and recall.
2) *Handling Negations and Sarcasm:* Implement strategies to detect and handle negations and sarcastic expressions in reviews, which can influence sentiment.

Text post-processing ensures that the sentiment analysis output is presented in a user-friendly and meaningful way. It bridges the gap between the raw model predictions and the information that end-users can readily interpret and act upon, making it a crucial step in the sentiment analysis workflow.

G. Deployment:

Deployment in sentiment analysis refers to the process of implementing and using sentiment analysis models or systems in a real-world or production environment. It involves taking a sentiment analysis model that has been developed and tested and making it available to analyze new, unseen data. Here's a detailed explanation of the deployment process in sentiment analysis:

1) *API Integration:* Set up an API that allows the e-commerce website to send new reviews for real-time sentiment analysis.
2) *Model Maintenance:* Continuously monitor and update the sentiment analysis model to adapt to changing language patterns and improve accuracy.

Deployment in sentiment analysis enables automated analysis of text data for sentiment classification, sentiment tracking, and analysis, and market research.

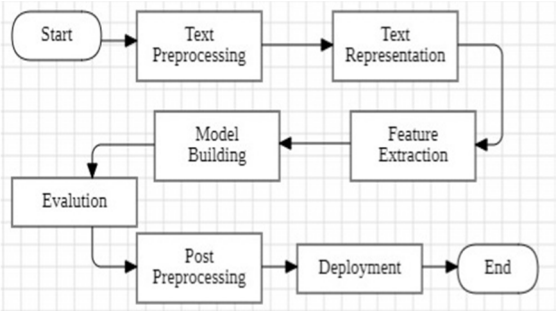


Fig.1 Process Flow

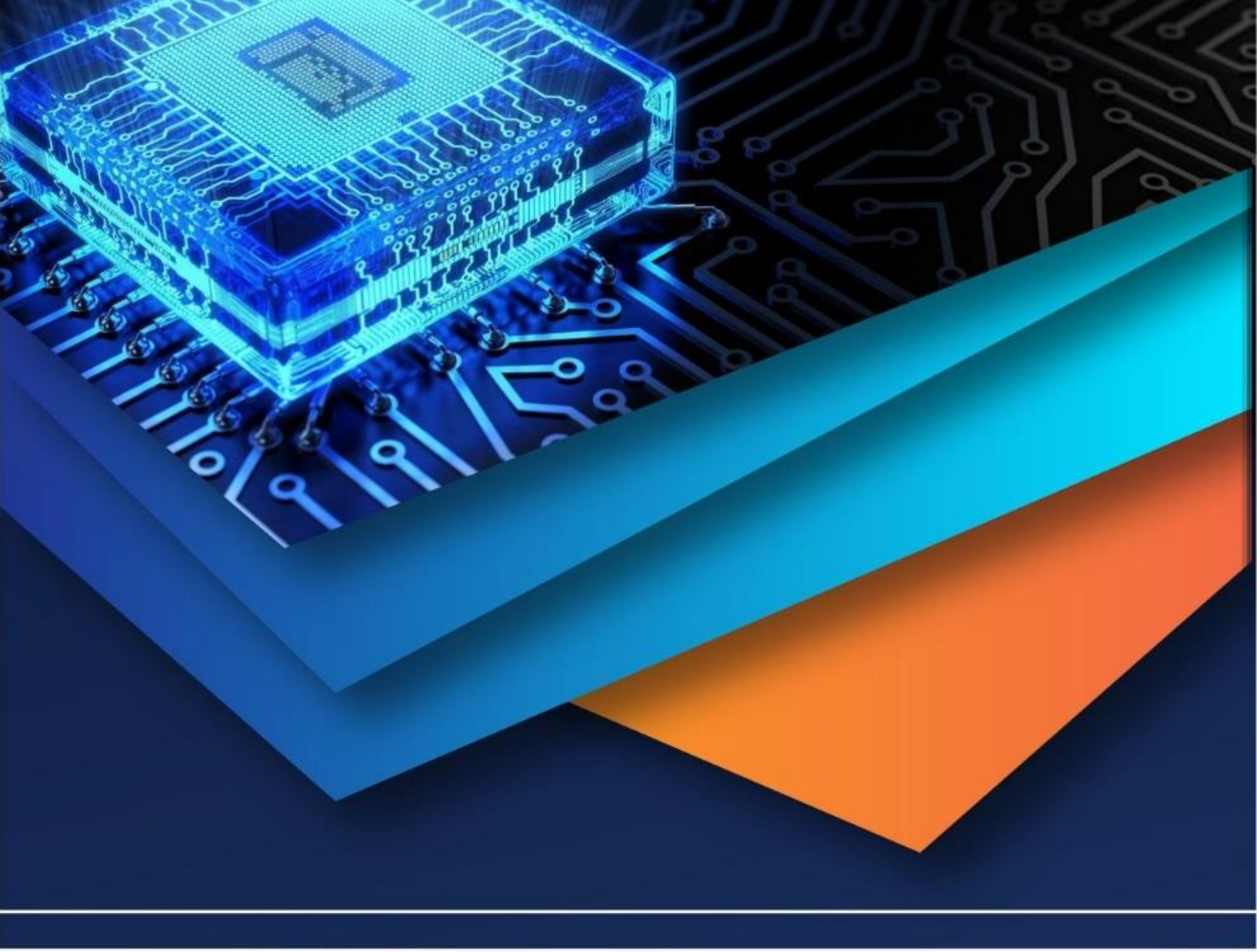
IV. CONCLUSION

Ratings and review summaries serve as a swift yet often speculative means of critiquing products. However, true critics require more comprehensive insights into customer experiences, which are typically found within the detailed product reviews. Reading through all these reviews can be quite inefficient. Therefore, this paper introduces a more precise and equally efficient approach utilizing Word Clouds and Sentiment Analysis. Word Clouds effectively highlight positive and negative aspects of a specific product. The attributes of words within a Word Cloud, such as font size, color, and whether they convey positivity or negativity, offer a straightforward representation of the product. Additionally, these Word Clouds offer valuable insights to manufacturers and sellers, helping them understand what features are liked or disliked by customers. By combining General and Specific Approaches, this model proves to be a more accurate and refined method for comparing two products or specific product features. With further enhancements, such as improved data cleaning, Part-Of-Speech (POS) tagging, and N-gram techniques to represent phrases in Word Clouds, as well as the introduction of color variations in Word Clouds for enhanced meaning, this model has the potential to replace the conventional, less accurate review systems.

REFERENCES

[1] Guo C., Du Z. & Kou X. (2018). Products Ranking Through Aspect-Based Sentiment Analysis of Online Heterogeneous Reviews. Journal of Systems Science and Systems Engineering, 27(5), 542–558.
[2] Rekha & Singh W. (2017). Sentiment analysis of online mobile reviews.
[3] PayalYadav & DhatriPandya. (2017). SentiReview: Sentiment analysis based on text and emoticons. 467-472.
[4] Sankar H. & Subramaniyaswamy V. (2017). Investigating sentiment analysis using machine learning approach. 87-92.
[5] Yang Kai & Cai Yi & Huang Dongping & Li Jingnan & Zhou Zikai& Lei Xue. (2017). An effective hybrid model for opinion mining and sentiment analysis. 465-466.
[6] Hasan A., Moin S., Karim A., &Shamshirband S. (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. Mathematical and Computational Applications,23(1).

[7] Shahzad Qaiser, Nooraini Yusoff, Ramsha Ali, Muhammad Akmal Remli, Hasyiya Karimah Adli.(2021). A Comparison of Machine Learning Techniques for Sentiment Analysis,12(3).
[8] Sobia Wassana, Xi Chenb, Tian Shenc, Muhammad Waqard, NZ Jhanjhie.(2021). Amazon Product Sentiment Analysis using Machine Learning Techniques, 30(1), 695-703.
[9] Monika Negi, Kanika Vishwakarma, Goldi Rawat, Priyanka Badhani. (2017). Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python, 165(9), 29-34.
[10] Rajalakshmi S. & Asha S. &Pazhaniraja N. (2017). A comprehensive survey on sentiment analysis. 1-5.
[11] V.K.Singh, R. Piryani, A. Uddin, P. Waila, "Sentiment Analysis of movie Reviews and Blog Posts", 3IEEE International Advanced Computing Conference (IACC), 2013
[12] Liu, Y., X. Hung, A. An, and X. Yu. Modeling and predicting the helpfulness of online reviews. In Proceedings of ICDM-2008,2008.
[13] Long, C., J. Zhang, and X. Zhut. A review selection approach for accurate feature rating estimation. In Proceedings of Coling 2010: Poster Volume, 2010.
[14] Lu, Y., P. Tsaparas, A. Ntoulas, and L, Polanyi, Exploiting Social context for review quality prediction. In Proceedings of International Word Wide Web Conference (WWW-2010),2010:AC.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)