



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Abstract Summarization of Product Reviews using Machine Learning Based on Sentiment Analysis

Prathamesh Mhetre, Amey Swami, Prajwal Thangan, Usha Jadhav, P. Malathi

Department of E&TC Engineering, D Y Patil College of Engineering, Akurdi, Pune, India

**ABSTRACT:** Customer testimonials are usually written in plain language and lack organisation. This data is organised using natural language processing techniques such as sentiment analysis are employed. This approach helps in determining whether reviews can be neutral, negative, or good. Sentiment analysis is instrumental in gauging the opinions and emotions associated with each product. The primary Predicting future product releases is the goal of the suggested research. To achieve this, machine learning algorithms are integrated with sentiment analysis to enhance performance. In this research, the initial step involves data collection and pre-processing. Subsequently, Sentiment analysis is utilized to evaluate customer reviews and extract relevant features. The performance is then optimized using Random Forest classifiers, followed by predicting future product releases through a decision tree algorithm. The proposed methodology shows improved performance outcomes compared to existing methods.

## I. INTRODUCTION

Potential customers frequently turn to online platforms to gather information about products, helping them decide whether to make a purchase. But going through every review that is out there might take time, and you can miss something important, potentially leading to poor purchasing decisions. Given that each product can have thousands of reviews, it becomes challenging for customers to make informed decisions based on the diverse opinions presented. Both positive and negative reviews exist for every product, complicating the decision-making process. Therefore, an overall product rating is often provided, based on individual features of the product. Customer feedback is crucial for enhancing product features.

To address these challenges, a more defined and concise method of presenting product reviews is proposed, allowing users to save time and effort by not having to read all reviews. Automatic review summarization is suggested as a solution to the ongoing issues customers face, using natural language processing (NLP) algorithms to structure the unstructured review data. Sentiment analysis is particularly useful for distinguishing between positive and negative feedback, playing a key role in text extraction. Future product predictions are made using decision tree techniques.

With numerous software products available in the market, choosing a specific product can be difficult. Based on online consumer reviews, the number of product users has changed in recent years. The dual presence of positive and negative reviews for each product complicates decision-making, necessitating an overall product rating based on individual features. Customer reviews are instrumental in product improvement.

To structure the unstructured review data for analysis, effective NLP algorithms are necessary. Sentiment analysis is preferred for evaluating positive and negative feedback. Unlike existing works that focus solely on extracting product reviews without implementing future predictions, this paper proposes an efficient approach by fusing machine learning methods with sentiment analysis. Hash vector functions are used to extract features from the analysed reviews, and Random Forest classification techniques are then applied to categorize reviews as positive, negative, or neutral. Future product predictions are made using decision tree techniques.

The format of the paper is as follows: The "Literature Survey," in Section 2, examines current methods. The approach, which includes data collection, data preprocessing, sentiment analysis for sentence extraction, feature extraction, sentiment classification, and future prediction, is described in Section 3, "Proposed Work." Section Four, 'ML Algorithms' describing the algorithms used in sentiment analysis and the classification process. Section 5, 'Experiment and Result,' assesses the obtained outcomes, and Section 6, 'Conclusion,' summarizes the research and discusses future work extensions.

## II. LITERATURE SURVEY

A paper [1] titled "Abstractive text summarization of reviews using Natural Language Processing" Authors proposed an NLP-based approach for abstractive text summarization of product reviews. Combining sentiment analysis and sequence-to-sequence models, they achieve concise summaries Food Reviews, the method enhances review analysis efficiency.

Paper [2] automated text summarization using deep learning and natural language processing was investigated in a survey. In this paper, authors discussed abstractive and extractive methods, focusing on challenges in maintaining context. The paper reviews deep learning models and suggests combining methods for improved accuracy.

Paper [3] The authors of "Automated Text Categorization Summarization using and Rule Reduction" described a rule-reduction-based approach for text categorization and summarization. Token Creation, Feature Identification, and Categorization/Summary are the three steps in their procedure. In document retrieval and metadata extraction. The method emphasizes interdisciplinary text mining.

Paper [4] The authors focused on sentiment analysis of feedback using a supervised machine learning approach. Using algorithms for supervised machine learning, with random forest, accuracy was highest in analysing data from 802 employees. The study highlights sentiment analysis's importance in understanding perceptions, with HR applications.

## III. PROPOSED WORK

Analysing product reviews is a widely accepted practice that helps users focus on their needs. However, previous research often neglects future predictions of software product specifications. To address this gap, this study proposes an approach that uses sentiment analysis to analyse software product specification reviews and a ranking system to forecast the product's future. Reviews are gathered only from input provided by customers. Six main processes make up the suggested methodology: Sentiment classification, feature extraction, sentiment analysis, sentiment analysis, data pre-processing, and future product prediction.

As seen in Figure 1, the architecture of the proposed methodology abstract Summarization of Product Reviews using Machine Learning Based on Sentiment Analysis.

**A Data Collection:** 34,000 user ratings from Kaggle are included in the dataset, with an emphasis on goods like Alexa, the Amazon Kindle app, and Stick Fire TV. The CSV format is used to store the dataset.

**B Data Pre-processing:** In this step, the review dataset is pre-processed using the following five steps: tokenization, HTML tag removal, stop word removal, accented character removal, and special character removal.

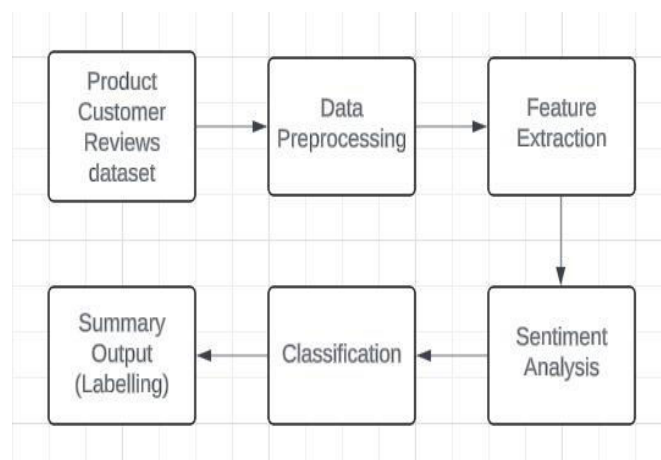


Fig.1. Architecture of the Abstract Summarization of Product Reviews using Machine Learning Based on Sentiment Analysis



Initially, HTML tags, hashtags, and usernames are removed as they do not contribute to text analysis. Tokenization follows, removing common, unnecessary words. The dataset is then cleaned by removing accented characters, stopwords, and special characters, making it ready for analysis.

**C Sentiment Analysis Sentence Extraction:** For sentiment analysis, objective content should ideally be removed, but here, it is retained for comprehensive analysis. Speech component (POS) Words are extracted via tagging, which keeps objective content, aiding in sentiment analysis. Table 1 shows a review with POS tags highlighted, and lists words with their corresponding POS tags. This process ensures that all review text is analysed without omitting objective content.

Table 1: Positive and Negative words for sentiment analysis

Positive Words	it!, turns!, blown, good!, wow!, down!, amazingly, book!, book!!, next!, read!, movie!, brilliantly, masterful, awesome, superb, fabulous, it!, wait, wonderfully, highly, turner!, incredible, toes, fantastic, bed, masterfully, thank, prime, loved, favour, I, blew, excellent, master, time!, chilling, amazing, crafted, end!, roller, story!, seat, loves, edge, gift, twice, beautiful, insightful, layers, constantly, wow, keeps, night, coaster, pieces, terrific, sleep, genius, predict, unpredictable, morning, thrilling, reading!, intricate, complex, fascinating, funny, immediately, enjoys, woven, late, unfolds, minute, love, beautifully, brilliant, surface, perfect, witty, till, fast-paced, intense
Negative Words	waste, poorly, wasted, worst, ridiculous, dumb, badly, awful, skipped, horrible, worse, depressing, pathetic, terrible, stupid, silly, boring, annoying, unrealistic, bother, poor, contrived, unbelievable, stuck, miserable, profanity, implausible, selfish, sorry, mistake, unlikeable, unlikeable, struggled, bothered, mess, quit, hated, death, book?, shallow, negative, disliked, cliché, dull, care, really?, annoyed, suspend, pass, sadly, cared, skip, holes, stopped, plain

**D Sentiment Scoring:** Sentiment scoring evaluates the emotions reflected in the text, assigning scores based on positive and negative counts. Reviews are scored on a scale from -1 (most negative) to +1 (most positive), with 0 being neutral. The method uses Vader sentiment analysis to determine the sentiment score and the ratio of sentiment scores given in equation (1) and (2). This analysis differentiates positive, negative, and neutral words, crucial for sentiment scoring. **Sentiment Score**= [(sum of positive words - sum of negative words) / (Total number of words in review)] --- (1)

**Ratio of sentiment Score**= (sum of positive words) / (sum of negative words + 1) --- (2)

**E Feature Extraction Sentiment scores** are extracted from raw datasets as features. Before classification, the data is converted into a feature vector, including sentiment tokens and hash values of strings. Each sentence is categorized with elements like hash values, positive and negative sentences, truth label, and average sentiment score. Values, both positive and negative, are calculated and shown in table 2:

Table 2: Examples of sentiment score

Consumer comment	Meaningful words	Sentiment weight	Sentiment score
Employees are always so polite and helpful. They have almost everything larger stores have, which is good for me! They even have white sweet potatoes which you usually only find at specialty grocers. Great experience!	Always Polite Helpful Good Great	1 2 2 1 2	8
Nice store. Seems clean and well organized. It is a bit on the smaller size as far as other stores go, so prepare for it to perhaps feel a little crowded, even if there aren't that many people in the store. This also means that their stock of certain items may be a bit smaller than the larger locations, but that's to be expected.	Nice Clean Well Organized Smaller Crowded Larger	2 2 1 2 -1(x2) -1 1	5
Since a new manager took over this once best supermarket around has fallen way off. They barely have anything, produce is always expired, and the store is not kept up very well.	New Best Fallen Barely Expired	1 2 -1 -1 2	-1

**F. Sentiment Classification:** The unbalanced data is classified to balance it using random forest classifiers, an approach to machine learning without supervision. Random forest classifiers are useful for both and improve prediction accuracy, issues with categorization and regression. Based on the bagging model, the dataset—which consists of numerous rows and columns—creates a number of decision tree models. The majority votes from each tree are added together to determine the final prediction.

**G. Future Prediction of the Product:** The classified and trained data is used to predict future product features. Decision trees, characterized by low bias and high variance, provide accurate results when fully constructed. The prediction results are combined, and majority vote decisions are made to convert high variance data into low variance, improving accuracy. Features with low variance are likely to be included in future product releases, while those with high variance may be removed or upgraded based on customer reviews.

#### IV. ML ALGORITHM

In this project, we implemented a Random Forest Classifier to analyse and summarize product reviews using sentiment analysis. We began by collecting and pre-processing the dataset, which involved cleaning and normalizing the text. Features were extracted using sentiment analysis techniques to create vectors representing the reviews. The Random Forest Classifier was then trained on the processed data, and its performance was evaluated using accuracy metrics and confusion matrices. Cross-validation and Grid Search were employed to optimize the model's hyper parameters, ensuring robust and reliable predictions. Finally, the sentiment analysis results were used to generate abstract summaries of the product reviews.

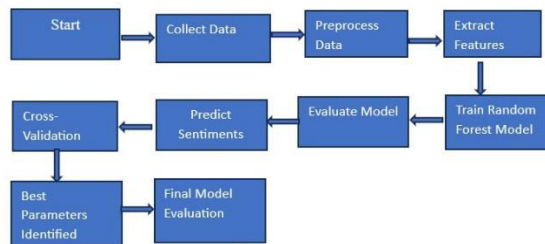


Fig. 2. Random Forest Classifier Flow Chart

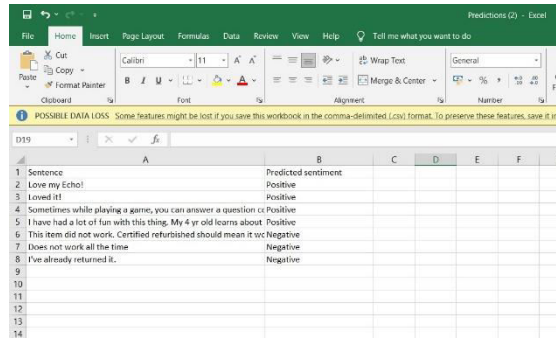
#### V. EXPERIMENT AND RESULT

This section provides a detailed a summary of the outcomes attained throughout the implementation stage. An examination of positive and negative reviews using word clouds was conducted. The word clouds, shown in Figures 3 categorize the reviews based on sentiment, aiding in the easy analysis of sentiments.



Fig. 3. Word Cloud of Positive Reviews

Figure 4 displays the results of positive reviews, listing the username, category, rating, and review text. The 'Sentence' column contains user reviews, the 'Predicted Sentiment' column specifies the either reviews are positive or negative. Similarly, tables for negative and neutral reviews were also evaluated.



	A	B	C	D	E	F	G
1	Sentence		Predicted sentiment				
2	Love my Echo!		Positive				
3	Loved it!		Positive				
4	Sometimes while playing a game, you can answer a question c		Positive				
5	I have had a lot of fun with this thing. My 4 yr old learns about		Positive				
6	This item did not work. Certified refurbished should mean it w		Negative				
7	Does not work all the time		Negative				
8	I've already returned it.		Negative				
9							
10							
11							
12							
13							
14							

Fig.4. Positive and Negative Review Result

Figure 5 illustrates the overall sentiment analysis results, categorizing reviews as positive, or negative represented by green and red colors, respectively. The analysis indicates that 57.1% of the reviews are positive, and 42.9% are negative. Figure 6 presents the product prediction results using a text. So, in this text view, it is observed that either product has positive or negative review. To forecast the products' future, a decision tree algorithm was used.

Text Sentiment Prediction

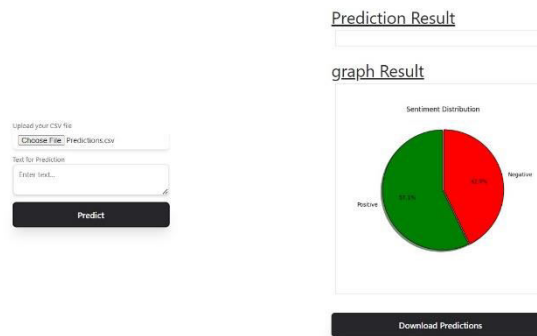


Fig. 5. Graph of Sentiment Analysis

Text Sentiment Prediction



Fig. 6. Result of Product Review Prediction

VI. CONCLUSION

This essay tackles the problem of projecting a product's future using user feedback. The suggested method illustrates how customers Sentiment analysis is used to examine reviews across a number of criteria. Additionally, emotion score is introduced, which is essential for feature extraction and categorization. Random Forest classifiers were used to improve classification performance, while decision tree algorithms were used to forecast product launches. The experiment visualizes the sentiment analysis results and future product forecasts. The outcome analysis reveals that 42.9% of the reviews are negative and 57.1% of the reviews are favourable. Subsequent research endeavours will centre on crafting a hybrid algorithm that demonstrates efficiency on several platforms.

#### REFERENCES

1. Dedhia, Parth & Pachgade, Hardik & Malani, Aditya & Raul, Nataasha & Naik, Meghana. (2020). Study on Abstractive Text Summarization Techniques. 1-8. 10.1109/ic-ETITE47903.2020.087.
2. S, Kavyashree & R, Sumukha & R, Soujanya & V, Tejaswini. (2023). Survey on Automatic Text Summarization using NLP and Deep Learning. 523-527. 10.1109/ICAECIS58353.2023.10170660.
3. Devasenal, C. & Hemalatha, M.. (2012). Automatic Text categorization and summarization using rule reduction. IEEE-International Conference on Advances in Engineering, Science and Management, ICAESM-2012.
4. H. A. Patrick, P. G. J, M. H. Sharief and U. Mukherjee, "Sentiment Analysis Perspective using Supervised Machine Learning Method," 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2023, pp. 1-4, doi: 10.1109/ICECCT56650.2023.10179807.
5. K. Dhola and M. Saradva, "A Comparative Evaluation of Traditional Machine Learning and Deep Learning Classification Techniques for Sentiment Analysis," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 932-936, doi: 10.1109/Confluence51648.2021.9377070.
6. T. L. Ben, N. Ravikumar R, P. C. R. Alla, G. Komala and K. Mishra, "Detecting Sentiment Polarities with Comparative Analysis of Machine Learning and Deep Learning Algorithms," 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 2023, pp. 186-190, doi: 10.1109/InCACCT57535.2023.10141741.
7. T. Velmurugan, M. Archana and U. Latha, "Sentiment Analysis of Customer Reviews Based Texts Using Classification Algorithms," 2022 8th Annual International Conference on Network and Information Systems for Computers (ICNISC), Hangzhou, China, 2022, pp. 185-191, doi: 10.1109/ICNISC57059.2022.00047.
8. S. Maurya and V. Pratap, "Sentiment Analysis on Amazon Product Reviews," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 236-240, doi: 10.1109/COM-IT-CON54601.2022.9850758.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details