

Sentimental Analysis of Product Review using Machine Learning



Swamini E Chavan, Archana D Ambhure, Mauli H Karche, Tushar A Kolhe, Vinita Kute

Abstract: Sentiment analysis of product reviews plays a crucial role in understanding consumer feedback, improving customer experience and making informed business decisions. This paper explores the application of machine learning and deep learning algorithms to effectively classify and analyse the sentiment of product reviews. Traditional machine learning techniques, such as Naïve Bayes, Support Vector Machines (SVM) and Random Forests are employed for sentiment classification based on manually engineered features. Simultaneously, deep learning approaches like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are leveraged to automatically learn complex representations from raw text data. The study compares the performance of these methods in terms of accuracy, precision, recall and F1-score. Additionally, pre-trained language models such as BERT are incorporated to enhance contextual understanding. Experimental results demonstrate that deep learning models particularly LSTM and BERT, outperform traditional machine learning techniques in capturing sentiments. This analysis provides valuable insights into the effectiveness of different algorithms in sentiment analysis tasks, paving the way for more advanced applications in natural language processing and customer sentiment evaluation.

Keywords: Sentimental Analysis, Deep learning, Naïve Bayes, Product Review, Feature Extraction.

I. INTRODUCTION

The rapid growth of e-commerce and online marketplaces has resulted in a substantial volume of customer-generated content, such as product reviews, ratings and feedback. This content serves as a valuable resource for understanding consumer preferences and behaviors, providing insights into customer satisfaction and product quality [1]. Sentiment analysis is also known as opinion mining, aims to extract

subjective information from text data, determining the sentiment expressed—positive, negative or neutral. It offers a systematic approach to analyze large volumes of textual data and derive meaningful insights, making it increasingly relevant for businesses, marketing strategies and customer service optimization [2].

Machine learning (ML) techniques have become a cornerstone in automating sentiment analysis due to their ability to learn from data and improve over time [3]. Traditional rule-based approaches, which rely on manually defined rules and lexicons, often struggle with the complexity and variability of human language [4]. In contrast, ML-based techniques leverage statistical models to capture the nuances of language and context, thereby enhancing the accuracy and scalability of sentiment analysis [5]. These models including supervised learning algorithms such as Support Vector Machines (SVM), Naive Bayes [6], Random Forests and deep learning architectures like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, have demonstrated significant potential in handling diverse datasets with varying degrees of sentiment complexity [7].

Sentiment analysis on product reviews not only benefits businesses by allowing them to monitor customer feedback in real-time but also empowers them to conduct competitive analysis [8], identify emerging market trends and personalize customer experiences [9]. For instance, detecting negative sentiments at an early stage can help companies take corrective actions to prevent customer churn [10]. Additionally, aggregating sentiment scores across product categories can facilitate the identification of strengths and weaknesses [11], enabling more informed decision-making in product development and marketing strategies [12].

II. SYSTEM OVERVIEW

The sentiment analysis framework focuses on its application in real-world scenarios, particularly for businesses and researchers. Several factors contribute in sentimental analysis are:

A. Automated Pre-processing

The framework automates essential pre-processing steps, such as text cleaning, tokenization and normalization. These steps address common challenges in natural language data, including handling punctuation, case sensitivity and stop words, thus reducing the need for manual data preparation. Users can quickly process raw review data with minimal configuration, lowering the barrier to entry.

B. User-Friendly Interface

To enhance accessibility, the system can be integrated with a graphical user interface

Manuscript received on 14 December 2024 | First Revised Manuscript received on 17 January 2025 | Second Revised Manuscript received on 30 January 2025 | Manuscript Accepted on 15 February 2025 | Manuscript published on 28 February 2025.

*Correspondence Author(s)

Swamini E Chavan, Bhivarabai Sawant Institute of Technology and Research (BSIOTR), Pune (Maharashtra), India. Email ID: swaminichavan285@gmail.com

Archana D Ambhure, Bhivarabai Sawant Institute of Technology and Research (BSIOTR), Pune (Maharashtra), India. Email ID: archanaambhure23@gmail.com

Mauli H Karche*, Bhivarabai Sawant Institute of Technology and Research (BSIOTR), Pune (Maharashtra), India. Email ID: maulikarche36@gmail.com

Tushar A Kolhe, Bhivarabai Sawant Institute of Technology and Research (BSIOTR), Pune (Maharashtra), India. Email ID: tkolhe2023@gmail.com

Prof. Vinita Kute, Bhivarabai Sawant Institute of Technology and Research (BSIOTR), Pune (Maharashtra), India. Email ID: vinitayedavekute@gmail.com

© The Authors. Published by Lattice Science Publication (LSP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

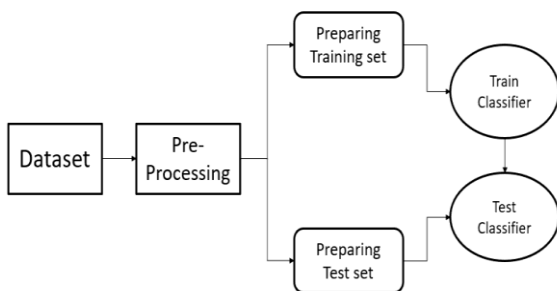


(GUI) that allows users to upload datasets, select analysis parameters and visualize results without writing code. Intuitive menus and tooltips guide users through the workflow making it suitable even for individuals with limited technical backgrounds.

C. Pre-trained Models and Transfer Learning

The framework incorporates pre-trained machine learning models such as BERT (Bidirectional Encoder Representations from Transformers) or Word2Vec, which have been fine-tuned on large corpora. These models reduce the need for extensive training and enable users to achieve high accuracy with smaller datasets. Transfer learning capabilities also allow for further fine-tuning based on domain-specific requirements.

III. PROPOSED SYSTEM



[Fig.1: Block Diagram of Sentimental Analysis]

A. Dataset

The dataset for sentiment analysis consists of product reviews, each labeled with a corresponding sentiment (e.g., positive, negative or neutral). It typically includes thousands of reviews collected from sources like Amazon, IMDb or a custom dataset.

Features:

- The primary feature is the text of the review.
- The label is the sentiment associated with the review (categorical for multi-class classification or binary for positive/negative classification).

B. Preprocessing Techniques

Text data needs to be processed into a format suitable for deep learning models.

Steps Involved:

- **Text Cleaning:** Remove punctuation, numbers, special characters and HTML tags to clean the text. Convert text to lowercase to ensure uniformity.
- **Tokenization:** Split the text into individual words or tokens. Tokenization can be done at the word or character level, though word-level is more common for sentiment analysis.
- **Stopword Removal:** Remove common words that may not add much value to the sentiment analysis (e.g., "the" and "is"). The NLTK library or custom stopwords lists can be used.
- **Stemming/Lemmatization:** Reduce words to their root form (e.g., "running" becomes "run"). Lemmatization is more context-aware than stemming.

- **Word Embedding:** Convert the text into numerical representations using techniques like Word2Vec, GloVe or embeddings provided by Keras. Embeddings transform words into vectors based on their contextual meaning.

C. Preparing the Training Set

- **Split the Data:** Typically, the dataset is split into 70-80% for training and 20-30% for testing. An additional validation set can be carved out from the training data (e.g., 10% of training data).
- **Data Augmentation:** Techniques like back-translation, synonym replacement and random deletion can be used to increase the size of the training set if needed.
- **Padding:** Deep learning models require inputs of fixed lengths. Pad the sequences to a uniform length using zero-padding, so that all review sequences have the same length.

D. Preparing the Testing Set

- The testing set should not be seen by the model during training.
- Preprocess the test set similarly to the training set (tokenization, padding, etc.).
- Keep the testing set separate to evaluate the generalization ability of the model after training.

IV. METHODOLOGY

Sentiment Classification Models

A. Machine Learning Models

- Naïve Bayes:** Naïve Bayes is mostly used as a baseline for text classification tasks. It calculates the probability of each sentiment class which is given the words in a review.
 - **Model Type:** It is Probabilistic classifier.
 - **Advantages:** Simple, fast, and performs surprisingly well on small datasets.
 - **Limitations:** It does not handle complex language structures.
- Logistic Regression:** Logistic regression is also popular baseline model for sentiment analysis. It uses weights for each word feature to predict the sentiment.
 - **Model Type:** It is Linear classifier.
 - **Advantages:** Easy to interpret and implement, works well with TF-IDF or BoW.
 - **Limitations:** It struggles with long-term dependencies in text.
- Support Vector Machines (SVM):** SVM is used for text classification tasks and can perform well with high-dimensional data such as TF-IDF.
 - **Model Type:** It is Linear classifier.
 - **Advantages:** It is good for medium-sized datasets which gives good performance.
 - **Limitations:** It can be slow with large datasets.

B. Deep Learning Models

- Convolutional Neural Networks (CNN):** Originally CNN is used for image processing but also effective in text analysis. It uses filters to detect local

patterns in text such as specific phrases or n-grams. It is suitable for capturing short-term dependencies in text. Fast to train due to parallel processing.

b. *Recurrent Neural Networks (RNN)*: RNN is designed for sequential data processing, making it suitable for text analysis. Maintains information from previous words (context) while processing the current word. It Struggles with long-term dependencies due to the vanishing gradient problem.

c. *Long Short-Term Memory (LSTM)*: A variant of RNN that addresses the vanishing gradient problem. Capable of learning long-term dependencies in text sequences. Uses memory cells to remember important information over long sequences. Widely used in sentiment analysis for handling longer text.

d. *Bidirectional Encoder Representations from Transformers (BERT)*: A transformer-based model that understands context from both left-to-right and right-to-left. Pre-trained on large datasets and can be fine-tuned for specific tasks (transfer learning). Achieves state-of-the-art results in various natural language processing tasks, including sentiment analysis. Effective at capturing complex language patterns and nuanced sentiments.

V. RESULTS AND DISCUSSION

Analyses are done on unstructured data collected from various resources. The system efficiency has been evaluated using precision, recall and F-measure which are as follows:

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall}$$

where, TP(True-positive), TN(True-negative), FP (False-positive) and FN (False-negative).

VI. CONCLUSION

User reviews are very important and they influence the purchase decisions. Sentiment analysis provides the users emotion towards the product and their services. Sentimental analysis can be implemented using various techniques and the results range accordingly to the conditions and the factors that influence them. In this paper, we have built and tested a model using logic on datasets of product reviews to find the sentiments of the reviews. The system performs well with the given dataset and with applied conditions

Deep learning significantly improves the accuracy of sentiment analysis in product reviews through several key ways such as automatic feature extraction, understanding context, handling large vocabulary and complex language. By using Machine Learning algorithms system can efficiently analyzes large volumes of data and provide valuable insights into customer opinions and attitude towards product.

DECLARATION STATEMENT

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted with objectivity and without any external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Authors Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. P. Rethina Sabapathi, Dr. K.P.Kaliyamurthie . "Analysis of Customer Review and Predicting Future Release of the Product using machine learning concepts". 2022 International Conference on Communication, Computing and Internet of Things. DOI: <https://doi.org/10.1109/IC3IOT53935.2022.9767956>
2. Shrey arora, Siddharth Mahapatra ,Dr. Anil Jadav ,Manojee Barla ,Nishant Mallick ." Temporal and Sentimental Analysis of Customer Reviews" . 2024 14th International Conference of Data Science. DOI: <https://doi.org/10.1109/Confluence60223.2024.10463473>
3. Yuvashree E, Preethika S, Nirupama A , Cloudin S ." Product Aspect Ranking Using Sentimental Analysis". 2021 International Conference on System, Computation, Automation and Networking. DOI: <https://doi.org/10.1109/ICSCAN53069.2021.9526429>
4. Gitanshu Chauhan, Akash Sharma, Nripendra Dwivedi. "Amazon Product Reviews Sentimental Analysis using Machine Learning ".2024 IEEE International Conference on Computing, Power and Communication Technologies. DOI: <https://doi.org/10.1109/IC2PCT60090.2024.10486258>
5. Hanan Alasmari." Sentimental Visualization: Semantic Analysis of Online Product Reviews Using Python and Tableau". 020 IEEE International Conference on Big Data (Big Data) | 978-1-7281-6251-5/20/\$31.00 ©2020 IEEE |. DOI: <https://doi.org/10.1109/BigData50022.2020.9391769>
6. Mukherjee, Subhabrata, and Pushpak Bhattacharyya. "Feature specific sentiment analysis for product reviews." International Conference on Intelligent Text Processing and Computational Linguistics. Springer, Berlin, Heidelberg, 2012. DOI: http://dx.doi.org/10.1007/978-3-642-28604-9_39
7. Somprasertsri, Gamgarn, and Pattarachai Lalitrojwong. "Extracting product features and opinions from product reviews using dependency analysis." 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery. Vol. 5. IEEE, 2010. DOI: <https://doi.org/10.1109/FSKD.2010.5569865>
8. Pathuri, S. K., Alapati, V. Ganesh., Teja, Ponnekanti. R., & Chowdary, R. (2019). Feature-Based Opinion Mining for Amazon Product's using MLT. In International Journal of Innovative Technology and Exploring Engineering (Vol. 8, Issue 11, pp. 4105–4109). DOI: <https://doi.org/10.35940/ijitee.k1837.0981119>
9. Sharma, P., & Site, S. (2022). A Comprehensive Study on Different Machine Learning Techniques to Predict Heart Disease. In Indian Journal of Artificial Intelligence and Neural Networking (Vol. 2, Issue 3, pp. 1–7). DOI: <https://doi.org/10.54105/ijainn.c1046.042322>
10. Rajasekaran, G., & Ram, Dr. C. S. (2023). Breast Cancer Prediction Based on Feature Extraction using Hybrid Methodologies. In International Journal of Soft Computing and Engineering (Vol. 13, Issue 2, pp. 20–28). DOI: <https://doi.org/10.35940/ijscce.b3612.0513223>
11. Harini, K., Yamuna, Dr. G., & Santhiya, V. (2020). Biometric Iris Recognition System using Multiscale Feature Extraction Method. In International Journal of Recent Technology and Engineering (IJRTE) (Vol. 8, Issue 6, pp. 2298–2303). DOI: <https://doi.org/10.35940/ijrte.f8016.038620>
12. Sharma, P. (2023). Advancements in OCR: A Deep Learning Algorithm for Enhanced Text Recognition. In International Journal of Inventive Engineering and

Sentimental Analysis of Product Review using Machine Learning

Sciences (Vol. 10, Issue 8, pp. 1–7). DOI:
<https://doi.org/10.35940/ijies.f4263.0810823>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Lattice Science Publication (LSP)/ journal and/ or the editor(s). The Lattice Science Publication (LSP)/ journal and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.