Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

Sentiment Analysis of Product Reviews using Rule-based and Deep-Learning Models

K.Surendra.[1], K.Nithin Prakash.[2], J.Maruthi Kumar.[3], G.Rakesh Goud.[4],N.Shanmugapriya[5]

[1-4]UG Student, [5]Associate Professor, AI&DS Dhanalakshmi Srinivasan University, Trichy, Tamil Nadu, India

Email: [1]kambellasurendra@gmail.com, [2]kunapareddynithin@gmail.com, [3]maruthikumarjogi224@gmail.com, [4]goragondllarakeshgoud@gmail.com

Abstract

This study analyses the feelings and opinions expressed in product reviews by utilizing NLTK (Natural Language Toolkit) and deep learning models to forecast customer sentiments and evaluate the probability of product purchases based on the review information available. The study examines the sentiment analysis applied to the Flipkart product reviews using a sentiment dataset. The NLTK, VADER, and RoBERTa models are evaluated for their effectiveness in predicting the sentiment of the customers. The analysis shows that the VADER, as a rule-based model is efficient in handling short and simple reviews but faces challenges with more complex sentiments. Meanwhile, RoBERTa outperforms VADER, with a Mean Absolute Error (MAE) of 0.12 and an R^2 value of 0.85. The comparative study shows the ability of RoBERTa to capture subtle emotions in customer reviews and accurately understand customer feedback, proving to be valuable in e-commerce for optimizing product recommendations and customer satisfaction.

Keywords: Sentiment Analysis, NLTK, VADER, RoBERTa, Transformers, Flipkart Reviews.

1. Introduction

Recently, social networking sites, such as Twitter, Facebook, Instagram, and many others have become popular platforms for sharing opinions about a particular incident, individuals, and events. The feedback offered is often constructive in helping to improve the service,

DOI: https://doi.org/10.36548/jtcsst.2024.3.007
Received: 11.08.2024, received in revised form: 12.09.2024, accepted: 30.09.2024, published: 15.10.2024

thereby enhancing customer satisfaction. Sentiment analysis is used to monitor and understand online reviews, enabling businesses to customize their products and services to meet customer needs and make the product successful. Figure 1 below shows the various types and algorithms of sentiment analysis [1,2].

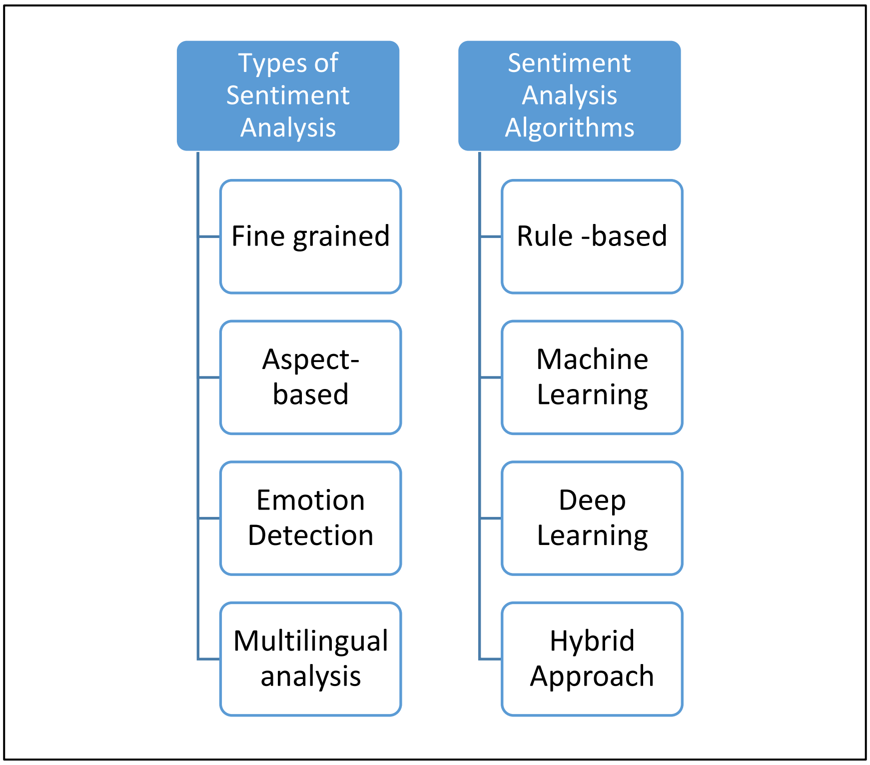

Figure 1. Types and Algorithms of Sentiment Analysis (https://marutitech.com/introduction-to-sentiment-analysis/)

Sentiment analysis has emerged as an essential tool in understanding customer opinions, particularly in e-commerce, where product reviews can significantly influence purchasing decisions. However, the field faces several challenges, including the complexity of natural language, the ambiguity in text, and variability in linguistic expressions. Traditional methods often struggle to capture the complex sentiment in large datasets, leading to inaccurate predictions. The study presents a comparison of the traditional rule-based model and the advanced deep learning model and highlights the superiority of the deep learning model over the traditional rule-based model [3-5].

The study aims to compare the performance of VADER, the rule-based model, and RoBERTa, an advanced deep learning model, using the sentiment dataset from Flipkart product reviews. The results show the ability of RoBERTa to optimize product recommendations and enhance customer satisfaction.

2. Literature Review

Sentiment analysis, a knowledge-based classification problem, is gaining popularity due to the growing number of reviews, surveys, and social media posts. The study discusses two AI-based approaches: rule-based reasoning and case-based reasoning, their principles, strengths, and limitations, and their application in sentiment analysis [6]. The research work in [7] suggests an approach to perform a party-based sentiment analysis using the rule-based approach and performs an aspect-based sentiment analysis to identify the level of sentiment in each sentence that is related to a particular party of a court case. The rule-based approaches, including Classic NLP techniques, stemming, tokenization, speech tagging, and machine learning parsing, using advanced Python language are adopted for the social media review comments of products and determines whether or not the text is positive, negative, or neutral [8]. Useful insights from sentiments can also be derived using machine learning approaches. The study in [9] compares the various machine-learning models and demonstrates, through the experimental analysis performed on the tweet dataset, that logistic regression outperforms other models support vector machine, naïve bayes with an accuracy of 82.5%. The author Verma and his team [10] conducted a survey on the sentiment analysis of social networking sites using lexicon and machine learning approaches to identify the opinions put forth by the users and determine the strength of the sentiments towards a particular product. This method uses lexicon resources to analyze reviews, handling polarity shifts, and incorporating intensifiers, punctuation, and acronyms. Words with opinions are extracted for scoring, and machine learning algorithms are applied. Experimental results show the model effectively identifies reviews' sentiments [11]. To overcome the limitations of the traditional approaches the deep learning approaches were introduced in the sentiment analysis. The review presents the recent studies on sentiment analysis on social networks, focusing on deep learning models to address challenges in natural language processing (NLP). It reviews models using the term frequency-inverse document frequency (TF-IDF) and word embedding and compares experimental results for different models and input features [12]. The deep learning algorithms, integrated with the machine learning models, perform better in analysing complex sentiments in large datasets and are utilized in organizations and businesses to make their products successful and enhance customer satisfaction [13]. The capability of the deep learning models to learn complex patterns automatically without manual feature extraction makes it a popular solution for sentiment analysis with limited labeled datasets [14]. This makes the deep learning algorithm well suited for the aspect-based sentiment analysis [15] and multimodal sentiment analysis [16], overcoming the challenges and enhancing the performance.

The proposed study compares the sentiment analysis using VADER [17] and RoBERTa [18] on the Flipkart review dataset and highlights the capability of the deep learning model over the rule-based approach in terms of Mean Absolute Error (MAE) and R-squared (R^2).

3. Methodology

The study compares two different sentiment analysis modules to analyze the customer review taken from Flipkart

Figure 2 below shows the flowchart of the comparative analysis.

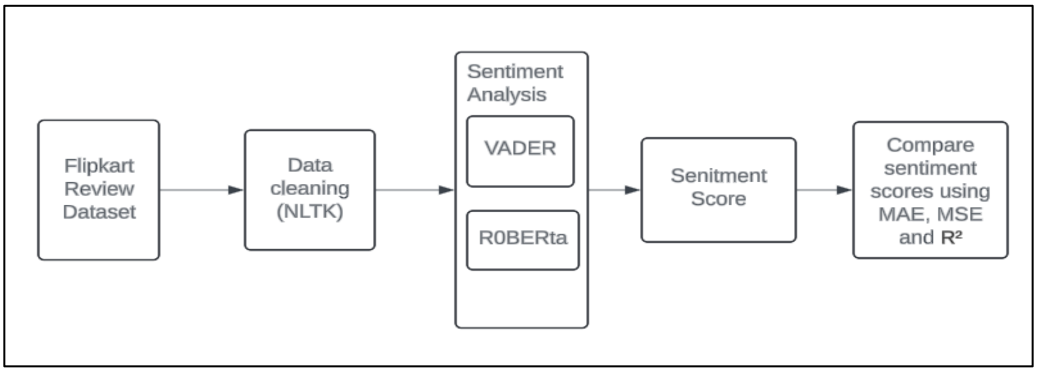

Figure 2. Flowchart of the Comparative Analysis

3.1 Data Collection

The proposed study utilizes the review comments for the products purchased by the customers from Flipkart, as dataset [19]. It includes 205,502 rows with main features of the product, such as product name, price, the ratings for the product, review comments for the product, and summary. The reviews were based on the experience of the customer about the 104 different product that are available on the Flipkart and the products purchased. The Figure 3 shows the review distribution for the star ratings.

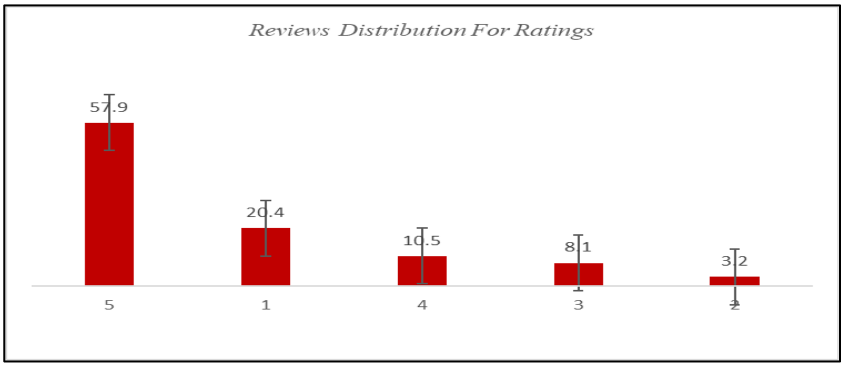

Figure 3. Review Distribution vs Star Ratings

3.2 Data Cleaning

The proposed study utilizes the NLTK (Natural Language ToolKit) to clean the dataset by removing the punctuation, special characters, and stopwords, as they do not carry significant meaning in sentiment analysis. It also performs tokenization, and lemmatization, and converts all the text to lowercase so as to reduce the redundancy and ensure consistency in the word representation.

3.3 Sentiment Analysis

The study compares the performance of VADER (Valence Aware Dictionary and sEntiment Reasoner) and the RoBERta on sentiment analysis using the Flipkart review dataset. VADER is a lexicon and rule-based sentiment analysis tool specifically designed for social media texts and short reviews. The pre-built sentiment lexicons are utilized to assign the sentiment scores to the individual words. Each tokenized and pre-processed review is analyzed using the VADER to obtain the sentiment score and classify the reviews as positive negative or neutral based on the compound scores: >0.05 (positive), -0.05 ≤ compound score ≤ 0.05(neutral), < -0.05 (negative). RoBERTa is a transformer-based model that is fine-tuned for the sentiment analysis task. The Hugging Face Transformers library is used in this implementation to enable efficient model configuration and training. Using the pre-processed Flipkart dataset, the RoBERTa model is optimized over five epochs, using a learning rate of 2e-5 and a batch size of 32. Cross-entropy loss is the selected loss function, and AdamW is the optimizer that is being used. The pre-processed reviews are sent into the RoBERTa model once it has been fine-tuned to produce sentiment predictions. This model thoroughly analyzes consumer sentiments in the dataset by using its contextual understanding of the text to generate sentiment categories. The Figure 4 shows the flow diagram of the sentiment analysis process.

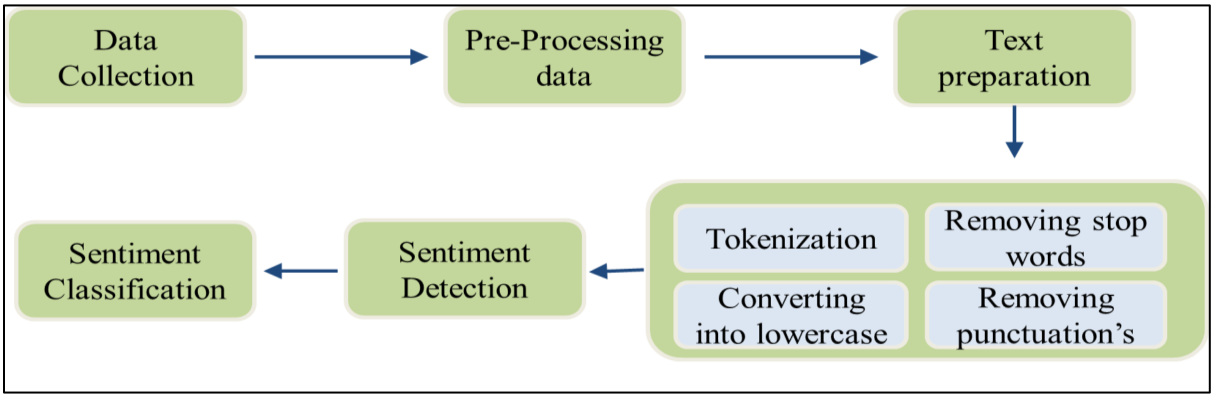

Figure 4. Module Flow Diagram

4. Results

VADER analyses the sentiment of the review without requiring training, as it is simply applied over the entire dataset. While RoBERTa the pre-trained transformer is fine-tuned using the 80% training dataset and evaluated on the 20% testing dataset.

Google Colab was utilized in the implementation of the proposed study. The libraries NLTK, Numpy, and Pandas were used for preparing data. VADER (from NLTK) was used for rule-based sentiment categorization in sentiment analysis, while RoBERTa was accessed through the Hugging Face Transformers library. PyTorch was necessary for model training and fine-tuning RoBERTa. Finally, Scikit-Learn was utilized to determine metrics including accuracy, precision, recall, and F1-score to evaluate the model's performance. Matplotlib and Seaborn were used to illustrate the results of the study.

4.1 Metrics Used

MAE—Mean Absolute Error: This was computed as the mean of the absolute differences between predicted sentiment scores $\hat{\boldsymbol{y}}_i$and the actual sentiment label $\boldsymbol{y}_i$. It states the average of how far the predictions are

$$MAE = \frac{1}{n}\sum_{i-1}^{n}|y_i - \hat{y}_i|$$

MSE—Mean Squared Error: This was calculated as the mean of the squared differences between the predicted sentiment scores $\hat{\boldsymbol{y}}_i$ and the actual sentiment label $\boldsymbol{y}_i$. MSE penalizes larger errors more heavily due to the squaring, making it useful for placing a stronger emphasis on larger deviations

$$MSE = \frac{1}{n}\ i = 1 \sum n(y_i - \hat{y}_i)^2$$

R-squared: This was calculated to assess how much variability in the sentiment data is explained by using a model. R-squared compares predicted sentiments in the model to a baseline model (the mean of the actual sentiments $\bar{\boldsymbol{y}}_{\boldsymbol{i}}$). Generally, a higher R^2 indicates a better fit of the model to the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

The Table 1 below shows the comparison of the evaluation metrics for the VADER and RoBERTa

Table 1. Performance Scores of VADER and ROBERTa

MODEL	MAE	MSE	R-SQUARED
VADER	0.18	0.05	0.70
ROBERTa	0.12	0.03	0.85

Mean Absolute Error (MAE) and Mean Squared Error (MSE) help determine how well the model's predictions align with the true values, and R^2 provides a measure of how well the variation in sentiment scores is explained by the model.

The MAE for RoBERTa is lower than that of VADER, indicating that RoBERTa's predictions are, on average, closer to the true sentiment scores than those of VADER. This suggests that RoBERTa provides more accurate sentiment predictions. Similarly, the MSE for RoBERTa is lower than that of VADER. This further supports the fact that RoBERTa's prediction errors are less frequent and less significant as MSE penalizes larger discrepancies more severely than MAE. Furthermore, RoBERTa's higher R^2 value indicates that it explains a larger percentage of the variance in sentiment scores compared to VADER. With an R^2 value of 0.85, RoBERTa captures 85% of the variability in sentiment scores, while VADER captures only 70%. This demonstrates that RoBERTa is more effective at modeling the underlying patterns in the data.

The Figure 5 and 6 shows the score assigned based on the review comments, such as "the product was good", I like the product", "Worst" and "Bad" etc, and the summary of the review features.

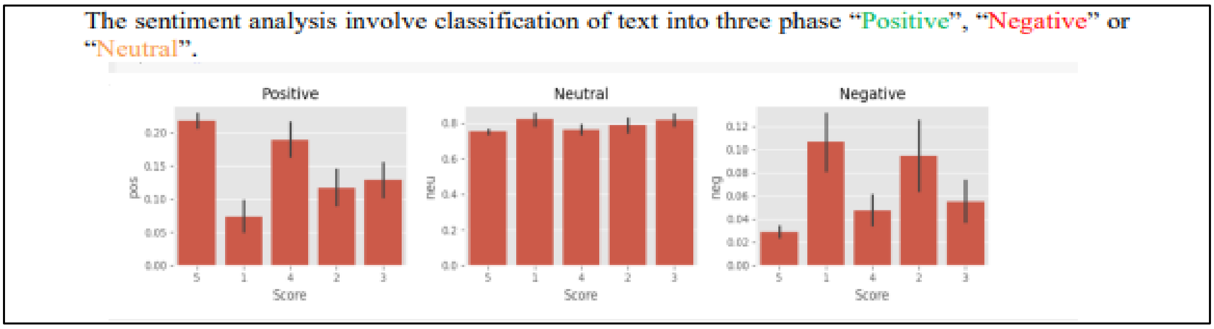

Figure 5. Sentiment Score for Rating Score

Figure 6. Summary of Review Features

5. Conclusion

The comparative study on the rule-based and the advanced deep learning model shows that RoBERTa performs better than VADER overall on all three metrics, outperforming VADER with a higher R-squared value and lower MAE and MSE values. These findings suggest that RoBERTa is a superior option for applications needing accurate sentiment predictions since it offers a more robust and accurate sentiment analysis than VADER. RoBERTa's improved performance can be due to its advanced deep learning architecture, which makes it more capable of comprehending language complexities and context than VADER's rule-based method.

References

[1] Aqlan, Ameen Abdullah Qaid, B. Manjula, and R. Lakshman Naik. "A study of sentiment analysis: Concepts, techniques, and challenges." In Proceedings of International Conference on Computational Intelligence and Data Engineering: Proceedings of ICCIDE 2018, Springer Singapore, 2019 pp. 147-162.

[2] Ray, Paramita, and Amlan Chakrabarti. "A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis." Applied Computing and Informatics 18, no. 1/2 (2022): 163-178.

[3] Kawade, Dipak R., and Kavita S. Oza. "Sentiment analysis: machine learning approach." International Journal of Engineering and Technology 9, no. 3 (2017): 2183-2186.

[4] Joyce, Brandon, and Jing Deng. "Sentiment analysis of tweets for the 2016 US presidential election." In 2017 ieee mit undergraduate research technology conference (urtc),Cambridge, MA, USA IEEE, 2017 pp. 1-4.

[5] Dadhich, Anjali, and Blessy Thankachan. "Sentiment analysis of amazon product reviews using hybrid rule-based approach." In Smart Systems: Innovations in Computing: Proceedings of SSIC 2021, Springer Singapore, 2022 pp. 173-193.

[6] Berka, Petr. "Sentiment analysis using rule-based and case-based reasoning." Journal of Intelligent Information Systems 55, no. 1 (2020): 51-66.

[7] Rajapaksha, Isanka, Chanika Ruchini Mudalige, Dilini Karunarathna, Nisansa de Silva, Gathika Rathnayaka, and Amal Shehan Perera. "Rule-based approach for party-based sentiment analysis in legal opinion texts." In 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer), Colombo, Sri Lanka, IEEE, 2020 pp. 284-285.

[8] Mitra, Ayushi, and Sanjukta Mohanty. "Sentiment analysis using machine learning approaches." Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2 (2020): 63-68.

[9] Nigam, Sandeep, Ajit Kumar Das, and Rakesh Chandra. "Machine learning based approach to sentiment analysis." In 2018 International Conference on advances in computing, communication Control and networking (ICACCCN),Greater Noida, India IEEE, 2018 pp. 157-161. .

[10] Verma, Binita, and Ramjeevan Singh Thakur. "Sentiment analysis using lexicon and machine learning-based approaches: A survey." In Proceedings of International Conference on Recent Advancement on Computer and Communication: ICRAC 2017, Springer Singapore, 2018 pp. 441-447. .

[11] Sahu, Tirath Prasad, and Sarang Khandekar. "A machine learning-based lexicon approach for sentiment analysis." In Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines, IGI Global, 2022 pp. 836-851. .

[12] Dang, Nhan Cach, María N. Moreno-García, and Fernando De la Prieta. "Sentiment analysis based on deep learning: A comparative study." Electronics 9, no. 3 (2020): 483.

[13] Kaur, Jaspreet, and Brahmaleen Kaur Sidhu. "Sentiment analysis based on deep learning approaches." In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS),Madurai, India, IEEE, 2018 pp. 1496-1500. .

[14] Pathak, Ajeet Ram, Basant Agarwal, Manjusha Pandey, and Siddharth Rautaray. "Application of deep learning approaches for sentiment analysis." Deep learning-based approaches for sentiment analysis (2020): 1-31.

[15] Wang, Jie, Bingxin Xu, and Yujie Zu. "Deep learning for aspect-based sentiment analysis." In 2021 international conference on machine learning and intelligent systems engineering (MLISE),Chongqing, China IEEE, 2021 pp. 267-271. .

[16] Ghorbanali, Alireza, and Mohammad Karim Sohrabi. "A comprehensive survey on deep learning-based approaches for multimodal sentiment analysis." Artificial Intelligence Review 56, no. Suppl 1 (2023): 1479-1512.

[17] Bhaumik, Ujjayanta, and Dharmveer Kumar Yadav. "Sentiment analysis using twitter." In Computational Intelligence and Machine Learning: Proceedings of the 7th International Conference on Advanced Computing, Networking, and Informatics (ICACNI 2019), Springer Singapore, 2021 pp. 59-66. .

[18] Briskilal, J., and C. N. Subalalitha. "An ensemble model for classifying idioms and literal texts using BERT and RoBERTa." Information Processing & Management 59, no. 1 (2022): 102756.

[19] https://www.kaggle.com/datasets/niraliivaghani/flipkart-product-customer-reviews-dataset