



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Sentiment Analysis of Amazon Reviews

Audhithiyah S, Durga N, Jerunez J, Aiswarya S

UG Student, Department of IT, Sri Venkateswara College of Engineering, Pennalur, Sriperumbudur, India

UG Student, Department of IT, Sri Venkateswara College of Engineering, Pennalur, Sriperumbudur, India

UG Student, Department of IT, Sri Venkateswara College of Engineering, Pennalur, Sriperumbudur, India

Associate Professor, Department of IT Sri Venkateswara College of Engineering, Pennalur, Sriperumbudur, India

ABSTRACT: In the contemporary digital landscape, online reviews wield considerable influence over consumer decision-making processes, and Amazon stands out as a global e-commerce giant hoisting an extensive reservoir of product evaluations. As consumers increasingly turn to online platforms for guidance, understanding the sentiments conveyed in these reviews become paramount. Sentiment analysis, a subfield of natural language processing extracting meaningful insights from the vast expanse of user-generated content on Amazon. This research endeavors to offer a comprehensive exploration of sentiment analysis techniques specifically tailored for Amazon reviews. The primary focus lies in the nuanced extraction and interpretation of sentiments expressed by customers within these reviews. By delving into the intricacies of sentiment analysis applied to the multifaceted realm of Amazon product evaluations, this study seeks to unravel the complexities inherent in user sentiments and contribute to a deeper understanding of the dynamics that shape consumer perceptions in the digital marketplace. The significance of this research extends beyond the realms of academia, as it holds the potential to inform businesses, marketers and even consumers about the intricate interplay of sentiments within the vast ecosystem of online reviews, ultimately influencing the decisions of a global consumer base navigating the expansive inventory of world's largest e-commerce platforms.

KEYWORDS: Valence Aware Dictionary for Sentiment Reasoning(VADER), Robustly Optimized BERT Pre-training Approach (RoBERTa), Natural Language Toolkit (NLTK), Exploratory Data Analysis (EDA), Long short-term memory (LSTM), CSV (Comma Separated Values) .

I. INTRODUCTION

The proposed research is dedicated to delving into the classification of sentiments derived from user reviews and ratings, with a particular focus on harnessing the extensive dataset of Amazon fine food reviews. This dataset serves as a rich source of information for exploring consumer sentiments and preferences within the realm of food products available on the Amazon platform. To effectively navigate through the complexities of the raw data, the Natural Language Toolkit (NLTK) is employed. This toolkit facilitates efficient tokenization, breaking down the textual data into meaningful units, and enables an insightful Exploratory Data Analysis (EDA), unveiling patterns and trends within the dataset. The sentiment classification process involves the utilization of two distinctive methodologies: the VADER sentiment analysis tool and RoBERTa. These approaches are chosen for their respective strengths in discerning sentiments and categorizing them into positive, neutral, and negative categories. By employing these methods, the aim is to unravel the nuanced spectrum of sentiments expressed within the Amazon fine food reviews, thereby offering valuable insights into user experiences and preferences. Furthermore, to enhance the visualization of the model's outcomes and make the findings more accessible, Seaborn pyplot is integrated into the research framework. This graphical representation not only allows for a comprehensive comparison of the results obtained from the VADER and RoBERTa models but also presents the findings in a structured and easily interpretable tabular format. Through the integration of these various tools and techniques, the proposed research endeavors to make a significant contribution to the field of sentiment analysis. It seeks to provide a detailed and insightful exploration of user sentiments within the context of Amazon fine food reviews, offering valuable insights for both consumers and food product vendors alike.

II. RELATED WORK

In [1] Authors explores The proposed approach uses the raw data which is translated into the necessary format during data preprocessing. Then, features from user evaluations are extracted using feature extraction algorithms which selects the best subset feature from the real feature list while maintaining the original data content. Finally, RNN- LSTM is used to classify sentiments to their suitable polarity. The outputs of the proposed system can be any one of these emotional labels: Very Positive, Positive, Neutral, Negative, or Very Negative.

In [2] Authors explores the paper which utilizes the Preferred Reporting Items for the Systematic Review and Meta-Analysis framework to report the findings on reviewing literature related to sentiment analysis in social media. It identifies challenges and other potential problems that scholars have encountered and suggests potential solutions.

In [3] the paper explains that we could still not find a comparable representation of advanced state-of-the-art techniques such as Transformers-based systems like BERT, T5, T0++, or GPT-2s. It uses several sentimental analysis techniques like Machine Learning Techniques , Lexicon Based Techniques.

In [4] This study uses TikTok reviews in English that were collected from Google Play Store. The data set consisted of 9000 review comments with varying star ratings. The extracted data, which includes the username, time, score, and review, is saved in CSV (Comma Separated Values) format. With all of these data variables, just the review text focus on this sentiment analysis.

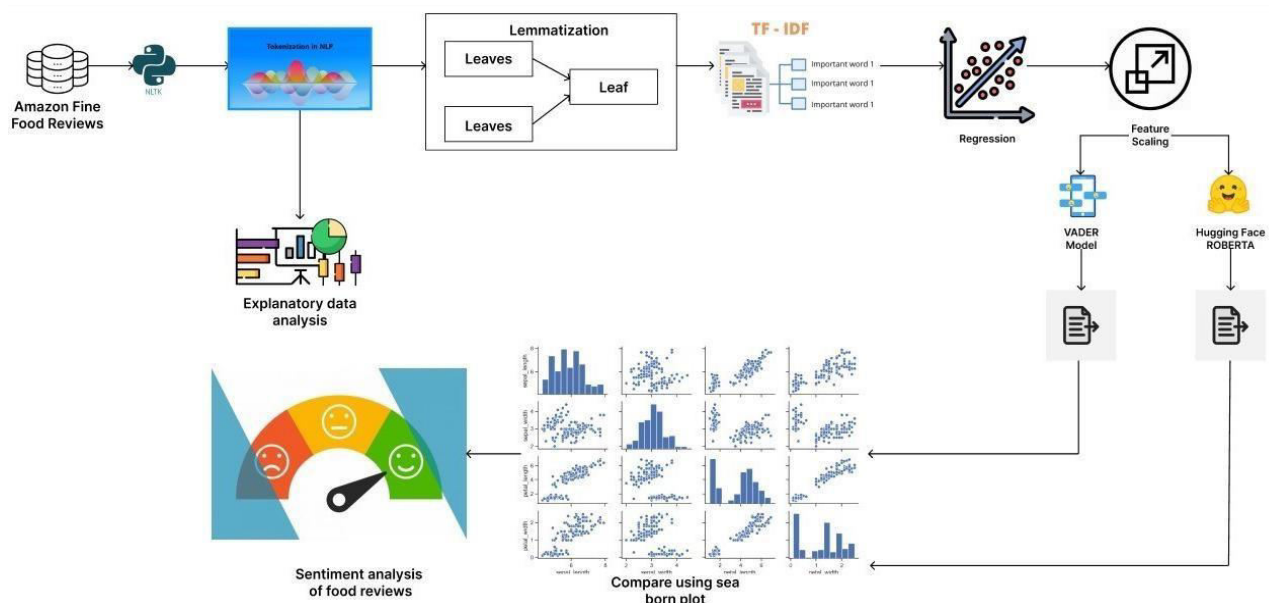
III. PROPOSED ALGORITHM

The proposed work involves using Amazon fine food reviews to analyze and compare the sentiments present in each review by employing two models: VADER and RoBERTa. The aim is to plot the sentiments together for comparison and examine their compound scores to understand the range of sentiments in the reviews, classifying them as positive, negative, or neutral.

The dataset used is Amazon_finefood_reviews.csv, which is 300.9MB in size and contains reviews of fine foods from Amazon. Spanning over a period of more than 10 years, it includes approximately 500,000 reviews up to October 2012. These reviews encompass product and user information, ratings, and plain text reviews, extending to reviews from all other Amazon categories as well.

The working methodology begins with the Natural Language Toolkit (NLTK), which is used to tokenize the raw data and perform an Exploratory Data Analysis (EDA). This step is crucial for breaking down the textual data into manageable pieces and uncovering initial patterns and insights within the dataset. Following the EDA, the sentiment classification is carried out using two advanced models: VADER and RoBERTa. VADER is specifically designed for sentiment analysis in social media texts, while RoBERTa is a robustly optimized BERT approach. Both models work to categorize the sentiments expressed in the reviews into positive, neutral, or negative.

To enhance the clarity and accessibility of the findings, the results are plotted using Seaborn pyplot. This allows for a comprehensive comparison of the sentiments identified by VADER and RoBERTa, presenting their compound scores in a visually appealing and structured format. By integrating these various tools and techniques, the proposed work aims to make a significant contribution to the field of sentiment analysis, offering a detailed and insightful exploration of user sentiments within the context of Amazon fine food reviews.



IV. PSEUDO CODE

```
# Run for Roberta Model
encoded_text = tokenizer(example, return_tensors='pt') output = model(**encoded_text)
scores = output[0][0].detach().numpy() scores = softmax(scores)
scores_dict = { 'roberta_neg': scores[0], 'roberta_neu': scores[1], 'roberta_pos': scores[2]
}
print(scores_dict)

fig, axs = plt.subplots(1, 3, figsize=(12, 3)) sns.barplot(data=vaders, x='Score', y='pos', ax=axs[0])
sns.barplot(data=vaders, x='Score', y='neu', ax=axs[1]) sns.barplot(data=vaders, x='Score', y='neg', ax=axs[2])
axs[0].set_title('Positive')
axs[1].set_title('Neutral') axs[2].set_title('Negative') plt.tight_layout() plt.show()
```

V. SIMULATION RESULTS

```
[51] #Positive 1-Star for VADER
      results_df.query('Score == 1') \
            .sort_values('vader_pos', ascending=False)['Text'].values[0]

      'So we cancelled the order. It was cancelled without any problem. That is a positive note...'

[53] #Negative 5-Star for ROBERTa
      results_df.query('Score == 5') \
            .sort_values('roberta_neg', ascending=False)['Text'].values[0]

      'this was sooooo deliscious but too bad i ate em too fast and gained 2 pds! my fault'

[55] #Negative 5-Star for ROBERTa
      results_df.query('Score == 5') \
            .sort_values('vader_neg', ascending=False)['Text'].values[0]

      'this was sooooo deliscious but too bad i ate em too fast and gained 2 pds! my fault'
```

VI. CONCLUSION AND FUTURE WORK

In conclusion, this project illuminates the pivotal role of sentiment analysis in navigating the contemporary digital marketplace, with a specific focus on the expansive realm of Amazon product reviews. The dominance of online reviews in shaping consumer decisions is undeniable, and Amazon, as a global e-commerce giant, plays a central role in hosting and disseminating these influential evaluations. The study delves into the intricacies of sentiment analysis, a powerful tool within the field of natural language processing, designed to distill meaningful insights from the vast sea of user-generated content on the platform. By concentrating on the nuanced extraction and interpretation of sentiments expressed by customers, the research provides a comprehensive exploration of the dynamics that underlie consumer perceptions in the digital landscape.

The significance of this work extends beyond academic realms, as its findings have practical implications for businesses, marketers, and consumers alike. Understanding the complex interplay of sentiments within Amazon's extensive ecosystem is essential for informed decision-making. The research underscores the potential of sentiment analysis to serve as a compass for businesses seeking to enhance their products or services, marketers aiming to craft effective strategies, and consumers navigating the intricate landscape of online reviews. As the digital marketplace

continues to evolve, this study contributes valuable insights that may guide global stakeholders in adapting to the ever-changing dynamics of consumer sentiments, ultimately influencing the choices of a diverse and expansive consumer base traversing the vast inventory of the world's largest e-commerce platform, Amazon.

REFERENCES

1. Rodríguez-Ibáñez, Margarita, Antonio Casáñez-Ventura, Félix Castejón-Mateos, and Pedro-Manuel Cuenca Jiménez. "A review on sentiment analysis from social media platforms." *Expert Systems with Applications* (2023): 119862.
2. Xu, Q. A., V. Chang, and C. Jayne. "A systematic review of social media-based sentiment analysis: Emerging trends and challenges." *Decision Analytics Journal*, 3, 100073. (2022).
3. Abiola, Odeyinka, Adebayo Abayomi-Alli, Oluwasefunmi Arogundade Tale, Sanjay Misra, and Olusola Abayomi-Alli. "Sentiment analysis of COVID-19 tweets from selected hashtags in Nigeria using VADER and Text Blob analyser." *Journal of Electrical Systems and Information Technology* 10, no. 1 (2023): 5.
4. Sangeetha, J., and U. Kumaran. "Sentiment analysis of amazon user reviews using a hybrid approach." *Measurement: Sensors* 27 (2023): 100790.
5. Isnan, Mahmud, Gregorius Natanael Elwirehardja, and Bens Pardamean. "Sentiment Analysis for TikTok Review Using VADER Sentiment and SVM Model." *Procedia Computer Science* 227 (2023): 168-175.
6. Kokab, Sayyida Tabinda, Sohail Asghar, and Shehneela Naz. "Transformer-based deep learning models for the sentiment analysis of social media data." *Array* 14 (2022): 100157.
7. Rakshitha, Kakuthota, H. M. Ramalingam, M. Pavithra, H. D. Advi, and Maithri Hegde. "Sentimental analysis of Indian regional languages on social media." *Global Transitions Proceedings* 2, no. 2 (2021): 414-420.
8. Abighail, Be Muhammad Doohan, Muhammad Reyhan Firmanda, and Maria Susan Anggreainy. "Sentiment Analysis E-commerce Review." *Procedia Computer Science* 227 (2023): 1039-1045.
9. Dhanalakshmi, P., G. Ashish Kumar, B. Sai Satwik, K. Sreeranga, A. Tharun Sai, and G. Jashwanth. "Sentiment Analysis Using VADER and Logistic Regression Techniques." In *2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS)*, pp. 139-144. IEEE, 2023.
10. Shahana, P. H., and Bini Omman. "Evaluation of features on sentimental analysis." *Procedia Computer Science* 46 (2015): 1585-1592.
11. Yue, Lin, Weitong Chen, Xue Li, Wanli Zuo, and Minghao Yin. "A survey of sentiment analysis in social media." *Knowledge and Information Systems* 60 (2019): 617-663.
12. Singh, Nikhil Kumar, Deepak Singh Tomar, and Arun Kumar Sangaiah. "Sentiment analysis: a review and comparative analysis over social media." *Journal of Ambient Intelligence and Humanized Computing* 11, no. 1 (2020): 97-117.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details