International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 25 Issue 02, FEB, 2025

RULE-BASED RECOMMENDATION SYSTEM FOR FOOD PRODUCT REVIEWS

Boina Naga Harsha vardhan[1], Dr. VSRK Sharma[2]
[1]*B. Tech Student, Dept. of Computer Science and Engineering, KL University, Vaddeswaram, Andhra Pradesh, India*
[2]*Deputy HoD, Dept. of Computer Science and Engineering, KL University, Vaddeswaram, Andhra Pradesh, India*

ABSTRACT:
An essential part of contemporary corporate intelligence and decision-making is the Amazon Food Product Reviews recommendation system. In a time when consumers frequently voice their thoughts online, it is critical for companies looking to succeed in cutthroat industries to comprehend and capitalize on the feelings expressed in customer reviews. Sentiment analysis is the study of how people's emotions are reflected. Emotions are the foundation of today's world. People use various behaviours to communicate their joy, sorrow, love, hate, and other emotions. Sentiment analysis is the division of emotions into positive, neutral, and negative categories. These days, sentiment-rich data can be found in tweets, status updates, blog entries, reviews, comments, discussion forums, and more. Analysing a website's rating entails looking at user reviews and ratings for the goods and services the platform offers. This analysis can offer insightful information on areas for improvement, product quality, and customer happiness. The data is gathered and the ratings are combined in this analysis. The shortcomings of the current system include continuous monitoring, effective analysis, and other issues. This study's main goal is to investigate the various methods and tools used to glean insightful information from consumer reviews. We explore the fundamental methods of sentiment analysis, including machine learning algorithms and Natural Language Processing (NLP), which allow reviews to be categorized into positive, negative, or neutral sentiments for efficient analysis. Finding the polarity of a particle review—whether it is poor, average, or outstanding for the various datasets—is the final step in drawing this conclusion.

I. INTRODUCTION

Sentiment analysis, also known as opinion mining, is a technique used to assess and summarize customer opinions and feedback from a variety of sources, including websites, e-commerce platforms, and social media. This method leverages linguistic, natural language processing (NLP), and text analysis methodologies to gauge the overall sentiment of a given piece of text, such as a review, comment, or opinion. Online sources, particularly e-commerce websites like Amazon, have become rich repositories of customer reviews and feedback on products. These platforms encourage customers to share their opinions about products and their shopping experiences. For instance, Amazon has built a vast database of customer reviews across multiple product categories, making it a popular resource for consumers seeking information about potential purchases. Consumers today rely heavily on reviews and ratings from others before making purchasing decisions. They seek insights from other customers' experiences to evaluate a product's quality and suitability for their needs. Businesses, in turn, can benefit from analyzing customer feedback to improve their products and services and ensure customer satisfaction.

Sentiment analysis plays a crucial role in evaluating customer feedback. By categorizing reviews as positive, negative, or neutral, it helps businesses understand customer opinions and identify areas for improvement. Moreover, it can guide customers in making informed decisions about products. In the case of Amazon's fine food reviews dataset, which contains over 500,000 reviews from users, sentiment analysis can be applied to classify reviews based on their polarity.

Using NLP and machine learning algorithms, the analysis can determine whether a review is poor, average, or excellent, helping businesses and customers gain valuable insights. As technology evolves, the ability to extract and interpret large volumes of data becomes increasingly important. Sentiment analysis offers a way to process unstructured data, enabling businesses and individuals to make data-driven decisions based on the sentiments expressed in customer reviews and feedback.

II. LITERATURE REVIEW

A literature review serves as a vital tool in academic and research endeavours, offering a thorough and critical summary of existing studies and scholarly articles focused on a specific topic or research question. It fulfils multiple roles within academic circles, aiding in the advancement of new knowledge, pinpointing gaps in existing research, and establishing a theoretical framework for subsequent studies. Usually presented as a standalone section in academic papers, dissertations, or theses, a literature review provides essential context, underlines the research's significance, and outlines the current understanding of the subject matter.

The process of conducting a literature review typically commences with the identification of the research topic or question, followed by a systematic exploration and selection of relevant literature from diverse sources like academic journals, books, conferences, and online databases. Researchers aim to gather a comprehensive array of perspectives and findings related to the chosen topic, ensuring that the review reflects the breadth and depth of scholarship in the field.

Once the relevant literature is amassed, the researcher critically evaluates and synthesizes the information, discerning key themes, trends, and debates. This rigorous evaluation involves scrutinizing the strengths and weaknesses of individual studies, evaluating the credibility and reliability of sources, and identifying prevalent patterns or discrepancies across the literature. Through this process, the literature review aims to offer insights into the current state of knowledge, highlight areas of agreement or disagreement among scholars, and identify gaps or unresolved questions that merit further exploration.

Literature survey about Sentiment Analysis Using Lexicon Based Approach:

A literature survey on sentiment analysis based on the lexicon approach involves examining research that utilizes dictionaries or lexicons to determine sentiment polarity in text data. This approach relies on predefined lists of words annotated with sentiment scores, which are used to assess the sentiment expressed in textual content. Researchers typically begin by compiling or selecting an appropriate lexicon tailored to the specific domain or language of interest. These lexicons contain entries for words along with their associated sentiment polarity, indicating whether they convey positive, negative, or neutral sentiment.

Studies in this field often explore various methods for sentiment analysis, including lexicon-based techniques such as the Bag-of-Words model or the Pointwise Mutual Information (PMI) approach. The Bag-of-Words model involves counting the occurrences of words from the lexicon within a given text and aggregating their sentiment scores to determine the overall sentiment. On the other hand, PMI calculates the association between words in the text and sentiment-bearing words in the lexicon to infer sentiment polarity.

Researchers also investigate the challenges and limitations of lexicon-based sentiment analysis, such as handling negation, sarcasm, context-dependency, and domain-specific language. Strategies for addressing these challenges may include incorporating linguistic rules, context-awareness, or machine learning techniques to enhance the accuracy and robustness of sentiment analysis systems.

Additionally, literature surveys often discuss applications of lexicon-based sentiment analysis across various domains, including social media analysis, product reviews, financial markets, and political discourse. Researchers examine how sentiment analysis can provide valuable insights for decision-making, opinion mining, brand monitoring, and sentiment tracking in real-time.

Overall, the literature survey on sentiment analysis based on the lexicon approach offers a comprehensive overview of the theoretical foundations, methodologies, applications, and challenges in this field. By synthesizing existing research findings and identifying emerging trends, researchers can contribute to advancing the state-of-the-art in sentiment analysis and

SNo	Tittle	Methodology	Advantages	Disadvantages
1	Sentiment Analysis Using Lexicon Based Approach	Lexicon-based approaches.	Simple and ease of implementation	Neutral Sentiment Challenge. Lack of Adaptability
2	Hybrid Approach for Sentiment Analysis	Combines multiple approaches like lexicon based and ML based methods	Improved accuracy compared to lexicon approach, better generalization	Increased storage. Increases Training Time. It effects the accuracy of overall analysis of reviews.
3	Sentiment Analysis Using Deep Learning	Deep learning techniques, LSTM, RNN	Preprocessing normalizes data for better analysis. LSTM provides advanced classification with memory cells and gates.	Requires expertise for manual sentiment determination
4	Sentimental Analysis and Prediction Using Neutral Networks	Sentiment analysis using TextBlob library Implementation of Artificial Neural Networks (ANN) using R Train and test dataset.	Reduced storage space requirements. Shorter training and testing time	Doesn't provide information on the hardware used for training the neural networks. Doesn't specify the specific version of TextBlob or other libraries used.

inform the development of more effective and versatile sentiment analysis techniques.

Challenges

Understanding people's feelings isn't always straightforward. Using machine learning to improve how we detect and analyze emotions, some hurdles, much like predicting and spotting potholes on reviews.

➢ **Neutral Sentiment Challenge:** Sentiment analysis encounters a significant hurdle when confronted with sentiments that neither lean overtly positive nor negative. These neutral sentiments often manifest subtly, are context-dependent, or possess nuanced characteristics, rendering their accurate categorization a challenge. Misclassifying neutral sentiments can distort overall sentiment analysis results, potentially leading to incomplete or misleading conclusions.

➢ **Inaccuracy of Overall Analysis of Reviews:** At times, sentiment analysis tools overlook the broader context or oversimplify the analysis process. Consequently, the overall sentiment score may fail to authentically capture the nuances inherent in individual reviews. Such inaccuracies in overall sentiment analysis can impede effective decision-making based on aggregated reviews.

➢ **Lack of Adaptability:** Certain sentiment analysis models struggle to adapt to new languages, dialects, or evolving language usage patterns. They may falter when encountering unfamiliar terms or expressions, limiting their effectiveness across diverse content and contexts.

➢ **Content-Dependent Errors:** The accuracy of sentiment analysis is heavily contingent upon the content under examination. Variations in writing style, tone, and contextual cues can introduce errors, resulting in incorrectly categorized sentiments. These errors can reverberate throughout downstream applications, impacting areas such as brand reputation management or product recommendations.

Addressing these challenges necessitates the development of robust models, context-aware algorithms, and a commitment to continuous improvement in sentiment analysis techniques.

III. PROPOSED SYSTEM

The proposed sentiment analysis system aims to improve upon existing approaches by incorporating advanced techniques and models for more accurate and reliable sentiment classification. This system includes the following key components and processes:

➢ **Data Collection:** The system begins with collecting a dataset containing text data from a relevant source, such as reviews or social media comments. Data sources should represent the specific domain or industry of interest to ensure that the sentiment analysis is applicable and meaningful.

➢ **Data Preprocessing:**

- **Noise Removal:** Special characters, punctuation marks, and unnecessary elements are removed to clean the data.
- **Text Lowercasing:** All text is converted to lowercase for uniformity and consistency. **Tokenization:** Text is split into individual words or tokens for further analysis. **Stemming/Lemmatization:** Words are reduced to their base form to avoid treating different inflected forms as distinct.
- **Stop word Removal:** Common stop words that do not carry significant meaning are removed.

➢ **Feature Extraction:**

- **Part-of-Speech Tagging**: Words are tagged with their grammatical categories to provide syntactic context.
- **Lexicon Analysis:** Words are mapped to sentiment lexicons that assign sentiment scores or polarity labels.
- **Syntactic Feature Extraction:** Dependencies and grammatical structures are analyzed to understand the relationships between words and phrases.

➢ **Model Training:**

- The system leverages advanced models such as RoBERTa for deep learning-based sentiment analysis and VADER for rule-based analysis.
- Training involves using labeled sentiment data to fine-tune the models for the specific domain.
- The system may use a hybrid approach combining rule-based and machine learning models for optimal performance.

➢ **Evaluation and Testing:**

- The system is tested using a portion of the dataset reserved for evaluation.
- Performance metrics such as accuracy, precision, recall, and F1-score are used to assess the effectiveness of the models.

- Iterative optimization is performed based on evaluation results to improve model performance.

➢ **Real-Time Analysis:** The system can be deployed for real-time sentiment analysis, providing immediate insights from incoming data. This allows businesses and other users to monitor sentiment trends as they happen and respond swiftly to feedback.

➢ **Actionable Insights:** The system provides users with detailed insights into sentiment trends, intensity, and patterns. These insights can be used to make data-driven decisions and tailor strategies based on customer feedback.

➢ **Scalability and Flexibility:** The system is designed to handle large volumes of data and scale as needed. It can be customized to various domains by adjusting training data and fine-tuning model parameters.

The proposed sentiment analysis system combines advanced preprocessing, feature extraction, and model training techniques to deliver accurate and insightful sentiment analysis. Its real-time capabilities, scalability, and actionable insights make it a valuable tool for businesses and other users to understand and act upon sentiment trends effectively.

Advantages

The proposed sentiment analysis system offers several advantages that make it a powerful and effective tool for understanding and acting on customer feedback and sentiment trends:

➢ **Accuracy:** By leveraging advanced models like RoBERTa and VADER, the system achieves high accuracy in sentiment classification. The combination of deep learning and rule-based approaches allows for capturing complex language nuances and contextual sentiment.

➢ **Real-Time Analysis:** The system can process and analyze incoming data in real-time, providing immediate insights into current sentiment trends. This enables quick responses to changes in customer feedback and emerging issues.

➢ **Actionable Insights:** The system offers detailed and actionable insights into sentiment trends, intensity, and patterns. These insights can guide strategic decisions, such as product development, marketing campaigns, and customer service improvements.

➢ **Customizability:** The system can be tailored to specific domains or industries by adjusting training data and fine-tuning model parameters. This flexibility allows for better performance in specialized areas such as healthcare, finance, or technology.

➢ **Scalability:** The system is designed to handle large volumes of data and can scale to accommodate growing datasets. Scalability ensures consistent performance even as data loads increase.

➢ **Enhanced Contextual Understanding:** Advanced models capture syntactic and semantic relationships between words and phrases, providing a deeper understanding of context. This enhances the ability to accurately classify sentiment even in complex sentences or reviews.

➢ **Integration Capabilities:** The system can be integrated with other data sources and platforms, such as social media channels or customer relationship management (CRM) systems. Integration enables a more comprehensive view of customer sentiment across different touchpoints.

➢ **Continuous Learning:** The system can be continuously trained and fine-tuned based on new data and feedback. This allows it to adapt to changes in language usage and emerging trends, maintaining high accuracy over time.

➢ **Cost-Effectiveness:** Automating sentiment analysis reduces the need for manual review and analysis, saving time and resources. By providing actionable insights, the system helps businesses optimize operations and improve customer satisfaction.

➢ **Increased Competitive Advantage:** The system provides businesses with timely insights into customer opinions, enabling them to stay ahead of competitors. Businesses can proactively address issues and capitalize on positive feedback to strengthen customer relationships.

Overall, the proposed sentiment analysis system offers a comprehensive solution for understanding customer sentiment, providing businesses with valuable insights and enabling them to make data-driven decisions for improved outcomes.

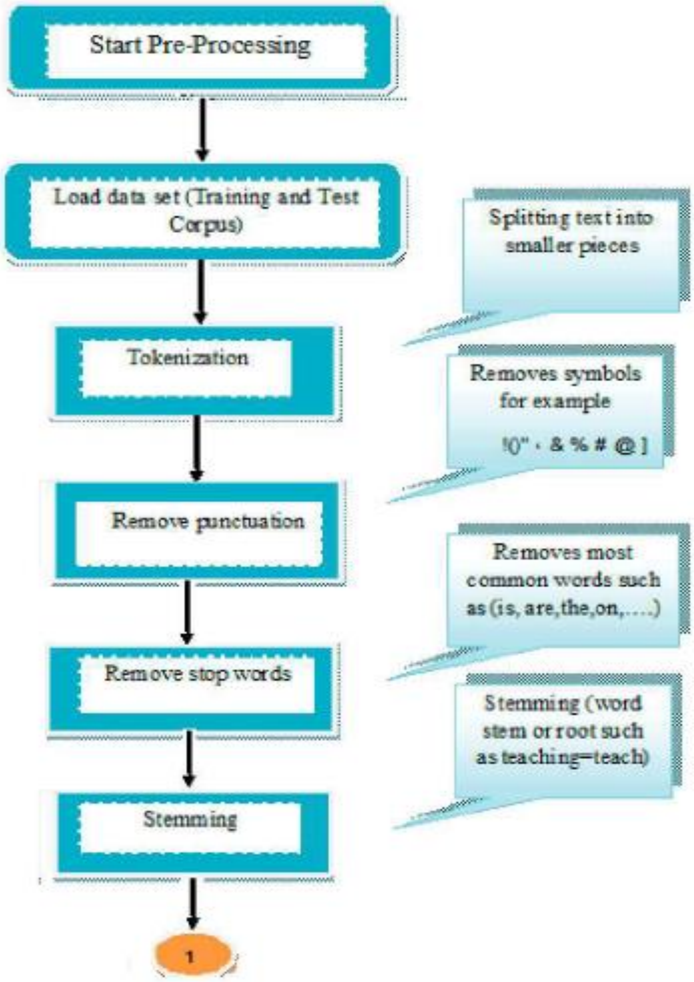


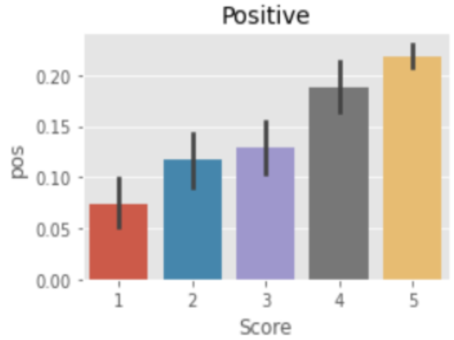
Fig: Proposed system architecture

IV. RESULTS

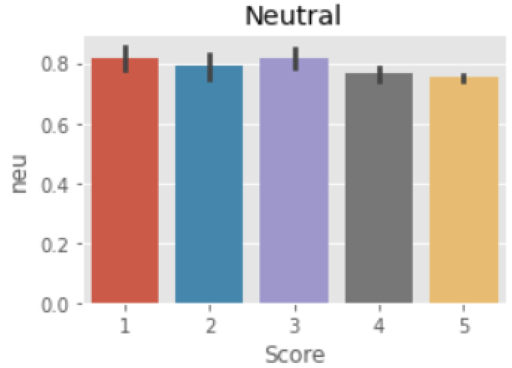
This is the output for count of reviews by stars

This is the output for Compound Score by Amazon Star Review

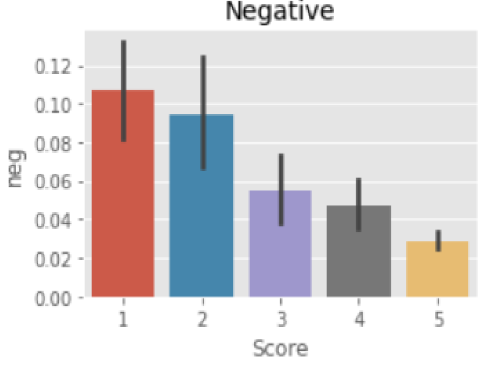
This is the output for Positive Compound Score by Amazon Star Review



This is the output for Neutral Compound Score by Amazon Star Review



This is the output for Negative Compound Score by Amazon Star Review



This is the code for count of reviews by stars

```
ax = df['Score'].value_counts().sort_index() \
    .plot(kind='pie',
          title='Count of Reviews by Stars',
          figsize=(10, 5))
ax.set_xlabel('Review Stars')
plt.show()
```

This is a code for Vanders Approach

```
sia.polarity_scores('I am so happy!')
```

```
{'neg': 0.0, 'neu': 0.318, 'pos': 0.682, 'compound': 0.6468}
```

```
sia.polarity_scores('This is the worst thing ever.')
```

```
{'neg': 0.451, 'neu': 0.549, 'pos': 0.0, 'compound': -0.6249}
```

This is the code for Roberta Approach

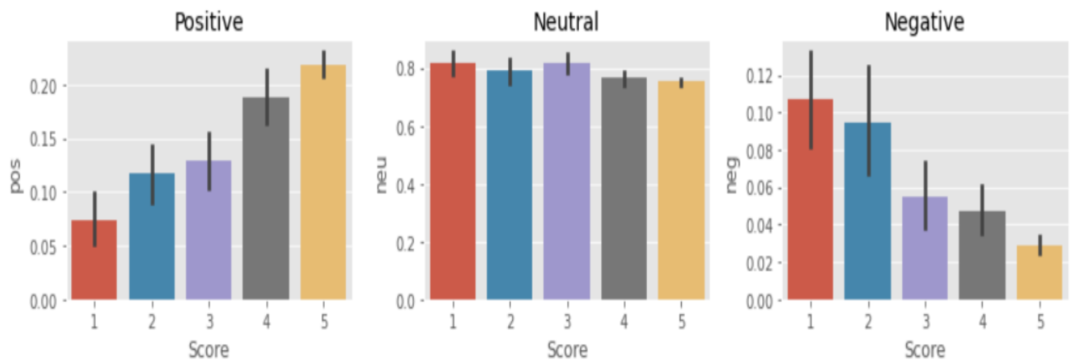
```
print(df['Text'][34])
sia.polarity_scores(df['Text'][34])
```

```
Instant oatmeal can become soggy the minute the water hits the bowl. McCann's Instant Oatmeal holds its texture, has excellent
flavor, and is good for you all at the same time. McCann's regular oat meal is excellent, too, but may take a bit longer to pre
pare than most have time for in the morning. This is the best instant brand I've ever eaten, and a very close second to the non
-instant variety.<br /><br />McCann's Instant Irish Oatmeal, Variety Pack of Regular, Apples & Cinnamon, and Maple & Brown Suga
r, 10-Count Boxes (Pack of 6)

{'neg': 0.0, 'neu': 0.874, 'pos': 0.126, 'compound': 0.9091}
```

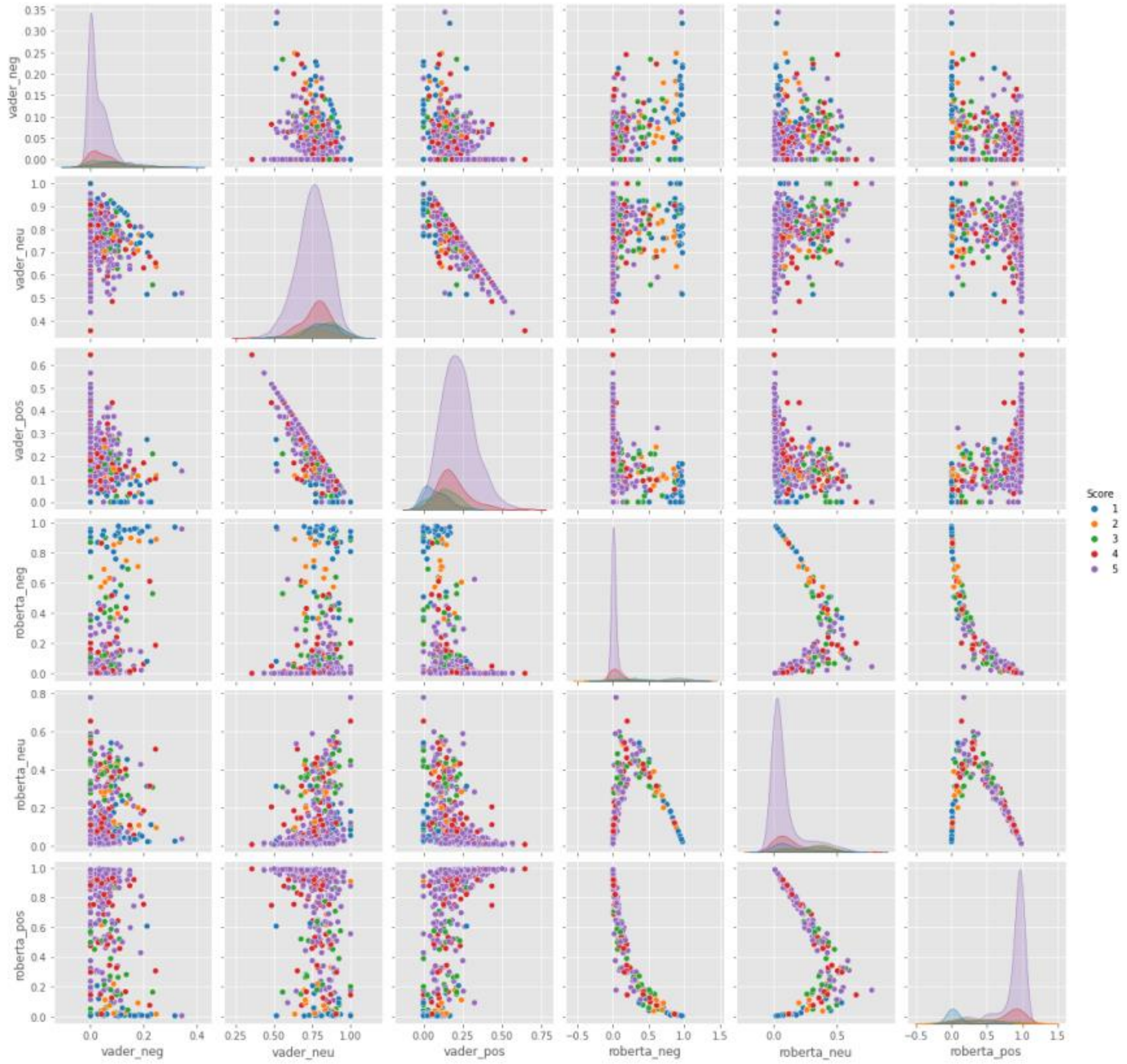
```
print(df['Text'][1])
sia.polarity_scores(df['Text'][1])
```

```
Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error
or if the vendor intended to represent the product as "Jumbo".

{'neg': 0.138, 'neu': 0.862, 'pos': 0.0, 'compound': -0.5664}
```

This is the output for all the Compound Scores by Amazon Star Review



This is the output for Difference between Vanders and Roberta Approaches



V. CONCLUSION

Sentiment analysis is commonly used approach to extract knowledge from text data in e-Commerce websites in the form of suggestions, feedback, and comments. By using sentimental analysis, we can analysis the customer emotion which helps the business to understand the customer requirement. The sentimental analysis is to identify the customer satisfaction and dissatisfaction on the product and help the merchants to improve their sales on the products. It also identifies the customer requirements in advance and help the ecommerce websites to develop their business.

The primary objective of this study is to explore the diverse approaches and technologies employed to extract valuable insights from customer feedback. We delve into the core techniques used in sentiment analysis, such as Natural Language Processing (NLP) and machine learning algorithms, which enable the classification of reviews into positive, negative, or neutral sentiments for effective analysis.

Finally, by this project we can conclude that our project is derived by finding the polarity of a particular review whether it is poor, average or excellent for the different datasets which helps E-commerce websites to predict the customer requirement in advance with the help of sentimental analysis using machine learning approach and helps to manage their product.

VI. REFERENCES

[1] Jiang, P., Zhang, C., Fu, H., Niu, Z., Yang, Q: An approach based on tree kernels for opinion mining of online product reviews. In: IEEE International Conference on Data Mining, pp. 256-265. IEEE Press (2010).

[2] Li, Z.: Product feature extraction with a combined approach. In: Third International Symposium on Intelligent Information Technology and Security Informatics, pp. 686-690. IEEE Press (2010).

[3] Adinarayana, S., Ilavarasan, E.: Classification techniques for sentiment discovery-a review. In: International Conference on Signal Processing, Communication, Power and Embedded System, pp. 396-400. IEEE Press (2016).

[4] Puri, S., Kaushik, S.: An enhanced fuzzy similarity-based concept mining model for text classification using feature clustering. In: Students' Conference on Engineering and Systems, pp. 1-6. IEEE Press (2012).

[5] Kaur, G., Singla, A.: Sentimental analysis of Flipkart reviews using Naïve Bayes and decision tree algorithm. International Journal of Advanced Research in Computer Engineering & Technology 5(1), 149-153 (2016).

[6] Khan, J., Jeong, B. S.: Summarizing customer review based on product feature and opinion. In: Proceedings of the 2016 International Conference on Machine Learning and Cybernetics, pp. 158-165. IEEE Press (2016).

[7] Puri, S., Singh, S. P.: A technical study and analysis of text classification techniques in N-lingual documents. In: International Conference on Computer Communication and Informatics, pp. 1-6. IEEE Press (2016).

[8] Kaur, J., Bansal, M.: Hierarchical sentiment analysis model for automatic review classification for E-commerce users. In: Banati H., Bhattacharyya S., Mani A., Köppen M. (eds) Hybrid Intelligence for Social Networks, pp. 249-267. Springer, Cham (2017).

[9] Ejaz, A., Turabee, Z., Rahim, M., Khoja, S.: Opinion mining approaches on Amazon product reviews: a comparative study. In: International Conference on Information and Communication Technologies, pp. 173-179. IEEE Press (2017).

[10] Tan, W., Wang, X., Xu, X.: Sentiment analysis for Amazon reviews. In: International Conference, pp. 1-5. (2018).