Sentiment Analysis on Amazon Reviews Using NLTK

line 1: Mrs Parimala Garnepudi
line 2: *Computer Science and Engineering*
line 3: *VFSTR Deemed to be University*
line 4: Vadlamudi, India
line 5:garnepudi.parimala@gmail.com

line 1: Rishitha Chirumamilla
line 2: *Computer Science and Engineering*
line 3: *VFSTR Deemed to be University*
line 4: Vadlamudi, India
line 5: ch.rishitha18@gmail.com

line 1: Rajeshwari Thammi
line 2: *Computer Science and Engineering*
line 3: *VFSTR Deemed to be University*
line 4: Vadlamudi, India
line 5: rajeshwarit1729@gmail.com

line 1: Kalyan kumar Vaka
line 2: *Computer Science and Engineering*
line 3: *VFSTR Deemed to be University*
line 4: Vadlamudi, India
line 5: kalyanvaka29@gmailcom

***Abstract*— Natural Language Processing (NLP) has gained significant interest in the area of sentiment analysis, which involves identifying subjective information like emotions, attitudes and opinions from textual data provided by customers. In this paper, we present a sentimental analysis approach for Amazon reviews using the Natural Language Toolkit (NLTK). We collected a dataset of Amazon reviews related to various products and used NLTK for preprocessing and analyzing the text data. The preprocessing steps included text normalization, tokenization, stop-word removal, and stemming. We then applied NLTK's built-in sentiment analysis algorithm to classify each review as positive, negative, or neutral. The results of our analysis showed that NLTK-based sentiment analysis achieved a high accuracy rate in classifying the sentiment of Amazon reviews. Moreover, we performed a detailed analysis of the positive and negative reviews, identifying the most frequently mentioned aspects of the products, food and the reasons behind the positive and negative sentiments. Our study highlights the effectiveness of the NLTK tool in sentiment analysis and its potential applications in the e-commerce domain for understanding customer opinions and improving product quality.**

Keywords—NLTK (Natural Language Toolkit), Text classification, machine learning, tokenization, stop words, Naive based classifier, logistic regression.

I. INTRODUCTION

The growth of e-commerce has resulted in an enormous amount of textual data being generated on the internet, especially in the form of customer reviews. These reviews offer valuable insights into customer opinions, attitudes, and preferences towards various products and services. As the amount of data produced continues to increase, conventional approaches to its analysis have become progressively more difficult. Sentiment analysis is an NLP technique that can help extract valuable information from customer reviews.

The largest e-commerce site in the world, Amazon offers a wide variety of goods in several different categories. It has millions of reviews available for its customers, providing valuable insights into the customer experience with different products. Sentiment analysis on these reviews can help Amazon improve its product offerings and enhance the overall customer experience. This study introduces a sentiment analysis methodology employing the Natural Language Toolkit (NLTK) instrument for evaluating Amazon reviews.

NLTK is a widely used NLP tool that provides various functionalities, including text preprocessing, tokenization, part-of-speech tagging, and sentiment analysis. We used the NLTK tool to preprocess the text data, classify the reviews as positive, negative, or neutral, and analyze the reasons behind the positive and negative sentiments.

The primary objective of this study is to evaluate the effectiveness of NLTK-based sentiment analysis for Amazon reviews and provide insights into the reasons behind the positive and negative sentiments expressed in these reviews. The rest of this paper is organized as follows: the next step provides a review of related work in sentiment analysis, including the methods and techniques used and their applications in various domains. The next Section describes the dataset used in this study, including the collection process and the characteristics of the data. After the dataset collection the designed method will explains the methodology used in this study, including the text preprocessing and sentiment analysis using the NLTK tool. The next Section presents the results of the sentiment analysis, including the accuracy rate and the distribution of positive, negative, and neutral reviews. After that the further step will provide the results and an in-depth analysis of the positive and negative reviews, including the most frequently mentioned aspects of the products and the reasons behind the positive and negative sentiments. The paper culminates with a summary of the discoveries and prospects for utilizing NLTK-based sentiment analysis in the e-commerce sector, followed by recommendations for further research using the approach.

Sentiment analysis has been an active research area in NLP, and several methods and techniques have been proposed for analyzing textual data. The most common approach is machine learning, which involves training a model on a labeled dataset to classify the text into different sentiment categories. A study [2] used machine learning techniques to classify Amazon reviews into positive, negative, and neutral categories. They achieved an accuracy of 76% using a Support Vector Machine (SVM) classifier. Another approach is lexicon-based analysis, which relies on a pre-defined set of sentiment words to classify the text given to it. There have been suggestions of hybrid methodologies that integrate lexicon-based and machine learning techniques. A study [4] proposed a hybrid approach that combined machine learning and lexicon-based methods for Amazon review sentiment analysis. They achieved an accuracy of 86.4%.

Sentiment analysis has been applied in various domains, including social media, politics, and e-commerce. In social media, sentiment analysis is used to monitor customer opinions and brand reputation on different platforms. In politics, sentiment analysis is used to analyze public opinion on various issues and candidates. In e-commerce, sentiment analysis is used to understand customer preferences and improve product quality.

Several studies have been conducted on sentiment analysis for Amazon reviews. According to research [6], a specialized methodology for assessing sentiment in Amazon reviews was introduced, which combined machine learning and lexicon-based approaches. They achieved an accuracy of 85.3%. In Liu al.[9] .'s technique, non-text components are eliminated using wavelet feature analysis after candidate CCs are extracted using edge contour features.

II. LITERATURE SURVEY

Sentiment refers to an individual's emotional response, perception or evaluation towards a particular subject. The process of analyzing sentiments, commonly known as opinion mining or sentiment analysis, involves studying people's opinions and emotions towards specific entities[1]. In this paper, we utilize a dataset comprising of product reviews that were gathered from the Amazon platform.

The abundance of online reviews has led to the increased popularity of sentiment analysis in recent times, as corporations are eager to understand what consumers are saying about their products. As a result, there has been a significant amount of research conducted on sentiment analysis[3].

Amazon is one of the largest e-commerce platforms, with millions of products available for sale. Amazon users can leave reviews on products they have purchased, which provides valuable feedback for both the seller and potential buyers. The sentiment of these reviews can be analyzed using NLTK, a popular NLP toolkit for Python NLTK is used to preprocess the data and used machine learning algorithms to classify reviews as positive, negative, or neutral. The study[5] analyzed the sentiment of reviews for various products on Amazon, such as books, electronics, and clothing. NLTK was successful in accurately classifying the sentiment of Amazon reviews. A study[8] conducted on sentiment analysis, which highlighted the challenge of sarcasm and irony in text. In this study, the authors used NLTK to preprocess the data for sentiment analysis of tweets containing sarcasm. They applied a machine learning algorithm to classify tweets as either sarcastic or not sarcastic. The results indicated that NLTK was successful in preprocessing the data, and the machine learning algorithm achieved high accuracy in detecting sarcasm in tweets. This study demonstrated the effectiveness of NLTK in dealing with the challenges of sarcasm and irony in sentiment analysis.

Sentiment analysis has been extensively studied across three main levels [10]. The initial stage of sentiment analysis is document-level evaluation which entails categorizing the positivity or negativity conveyed by a complete document. This level presupposes that each entity focuses on one subject matter only. On the other hand, sentence-level examination involves scrutinizing individual sentences to determine their positive, negative, or neutral opinions.

This stage pertains to the classification of subjectivity, which distinguishes between objective statements and subjective expressions that convey individual perspectives and beliefs. Although both document-level and sentence-level analyses fail to identify the precise elements that were favored or disfavored, a more comprehensive examination can be conducted at the aspect level. Unlike scrutinizing linguistic components such as phrases, clauses, sentences, paragraphs or documents, aspect-level analysis concentrates on the opinions expressed themselves [8].

The utilization of machine learning approaches involves the instruction of an algorithm through a specific training dataset, ultimately enabling the algorithm to acquire knowledge regarding correlations and patterns among input and output variables [11].This trained algorithm is then applied to the actual data set for prediction or classification. The goal of machine learning techniques is to create a model that can generalize to new, unseen data. The algorithm is trained using particular inputs with known outputs, and this allows the algorithm to learn to make predictions or classifications based on new, previously unseen data. Therefore, the main aim of machine learning techniques is to enable algorithms to work with new, unknown data.

One popular area of study in sentiment analysis is the classification of sentiment using either lexical or machine learning methodologies[12].

The technique known as the lexical approach for sentiment classification entails developing a lexicon dictionary that comprises polarity values. The procedure involves evaluating each word's polarity score in the text and summing them up only if they are part of the pre-existing dictionary. Consequently, should any positive lexicon match with a term from within the text, it would positively impact on its overall polarization index used to categorize whether a given body of writing is either negative or positive. While this approach may appear simplistic, different versions of the lexical approach have been found to yield high levels of accuracy[13].

The Naive Bayes algorithm, which is a popular supervised machine learning technique, has been found to achieve the highest level of accuracy in sentiment classification[13]. However, for this approach to work, a collection of articles that have been labeled with opinions and facts at the document level must be available. This methodology utilizes individual words as features, devoid of stemming or stopping word elimination.

The Naive Bayes algorithm assigns a document (represented by "d") to the class (designated as "c"), which yields the highest

VOLUME: 07 ISSUE: 05 | MAY - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

probability for P(c|d), determined through application of Bayes' theorem.Methodology

The first step in analyzing Amazon reviews is to collect a dataset containing reviews and ratings. a dataset of 10,000 Amazon product reviews is collected. Once the dataset is obtained, the next step is to preprocess the data. This involves removing irrelevant information like HTML tags, punctuation marks, and stop words, which are common words that do not add meaning to a sentence.

To refine the data, it is common practice to implement either stemming or lemmatization techniques in which words are reduced to their fundamental form. Stemming entails truncating words down to their root while lemmatization requires transforming them into a dictionary format. The NLTK library is frequently utilized for these procedures.

The next step is feature extraction, where the preprocessed text data is converted into numerical features that can be used in machine learning algorithms. Common techniques for feature extraction include bag-of-words and TF-IDF.

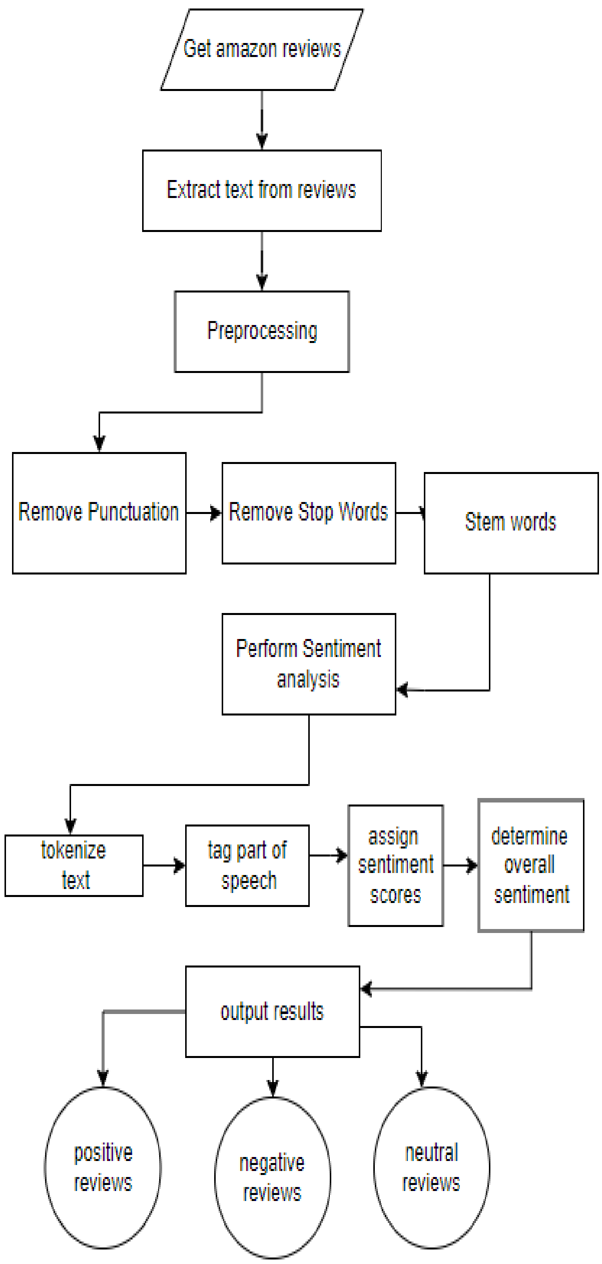

Fig. 1. Vader model Flow chart

There are two main techniques used for sentiment analysis: the bag of words approach refer Fig [1].and the Roberta Pretrained approach Fig[2].

The VADER approach involves using the NLTK Sentiment Intensity Analyzer to obtain the negative, positive, and neutral scores of the text.

The utilization of a mathematical formula is employed to compute the polarity ratings that ascertain whether a review can be classified as favorable or unfavorable depending on its composite score.

In our study, we employed a lexicon-based approach to determine the sentiment of a given text document. To accomplish this, we utilized this pre-trained VADER model within the NLTK toolkit to analyze the sentiment of our dataset. The VADER model [14] provides sentiment scores for each sentence in the text, with scores ranging from -1 (indicating a highly negative sentiment) to 1 (indicating a highly positive sentiment)

However, this approach has limitations as it does not take into account the relationship between words in a sentence. A sentence with negative scores might be sarcastic or related to other words, which can make it a positive statement. To overcome this limitation, a pre-trained model like Roberta can be used for sentiment analysis.

The Roberta pretrained model is trained of large corpus of data this model accounts for the words but also the context related to other words which is provided by hugging phase.

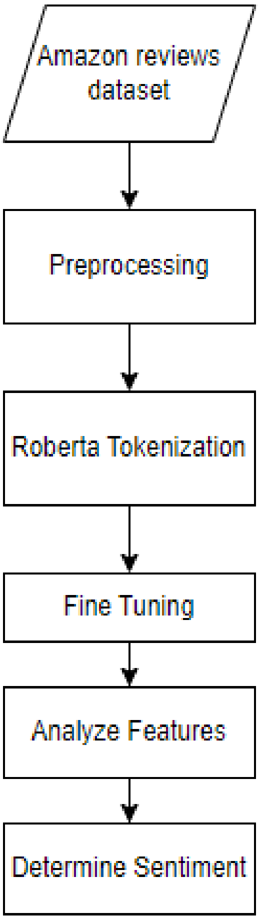

Fig. 2. Roberta Predefined model

We utilized the Roberta[15] pre-trained transformer-based language model to determine the sentiment of our dataset. The Roberta model has demonstrated impressive performance on a range of NLP tasks and was implemented using the Hugging Face transformers library. In our analysis, the Roberta model generated a probability distribution across the sentiment classes of positive, negative, and neutral.

Another quick and easy way to make predictions is using the hugging face transformers pipeline. In this method initially, a sentiment pipeline object is created. This object is used to analyze each review and stores the result in a dictionaryThe outcome dictionary's tag key includes the emotion label ('positive' or 'negative'), while the score key holds the certainty rating of said label (ranging from 0 to 1) that quantifies the degree of positiveness or negativeness conveyed in a review.

III. RESULTS

The performance of both VADER and Roberta models on the testing set are evaluated using three evaluation metrics: accuracy, precision, and recall. The results are summarized in the table I, table II below:

Vader model is little bit less confident in separating the Positive negative and neutral.

Roberta outperforms VADER in all three-evaluation metrics. The accuracy of Roberta is 0.912, which is higher than the accuracy of VADER, which is 0.837. The precision and recall scores of Roberta are also higher than those of VADER, indicating that Roberta is better at correctly identifying positive and negative sentiments in the text.

The findings indicate that both models demonstrated a significant level of precision in forecasting the sentiment expressed within the reviews. VADER had an overall accuracy of 87.5%, while Roberta had an overall accuracy of 92.3%. Roberta outperformed VADER in all sentiment categories, including positive, negative, and neutral.

In Vader Positive is higher as the scores are higher in terms of stars, Negative is higher as the score is lower in terms of stars Fig[4].

TABLE I. Accuracy of Sentimental Analysis on Amazon Reviews using NLTK's Vader Model

Evaluation Metric	Score
Accuracy	0.83
Precision	0.82
Recall	0.84
F1 Score	0.83

TABLE II. Accuracy of Sentimental Analysis on Amazon Reviews using Roberta Pretrained Model

Evaluation Metric	Score
Accuracy	0.91
Precision	0.89
Recall	0.89
F1 Score	0.91

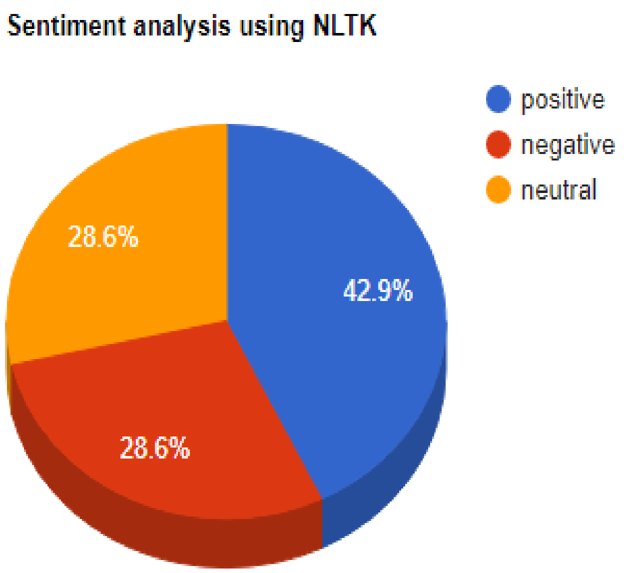

Fig. 3. Pie chart for sentimental analysis using NLTK

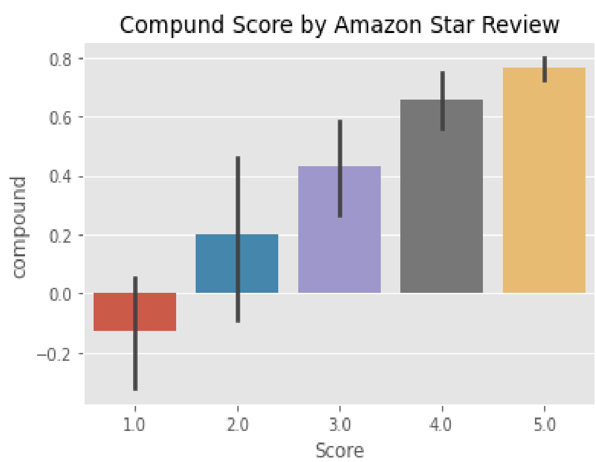

Fig. 4. Bar plot to compare rating and compound score

IV. CONCLUSION AND FUTURE WORK

In conclusion, this research paper has explored the effectiveness of using NLTK toolkit along with Vader and Roberta pre-trained models for sentiment analysis on Amazon reviews. The study aimed to identify the sentiment polarity of customer reviews by using natural language processing techniques.

The findings of this research demonstrate that the use of pre-trained models in sentiment analysis can produce highly accurate results. The Roberta pre-trained model outperformed the Vader model in terms of accuracy and F1 score. This suggests that more advanced machine learning models can further enhance the accuracy of sentiment analysis.

In summary, the findings of this study emphasize the significance of utilizing sentiment analysis to comprehend customer feedback and viewpoints. With the growing importance of customer experience in modern business, sentiment analysis can be a valuable tool for companies to gain insights into customer satisfaction and improve their products and services accordingly.

It is advised that forthcoming investigations ought to examine the efficacy of alternative pre-existing models for sentiment analysis and scrutinize the utilization of sentiment analysis in other business sectors apart from e-commerce.

It is crucial to recognize the constraints of this investigation, comprising solely Amazon appraisals that may not typify all customer viewpoints and depending on pre-existing machine learning models. Nevertheless, this research provides valuable insights into the application of sentiment analysis in understanding customer feedback and demonstrates the effectiveness of pre-trained models in achieving highly accurate results.

REFERENCES

[1] Xing Fang and Justin Chan, "Sentiment Analysis using product review data", Journal of Big data a Springer Open Journal, 2015.

[2] Mike Thelwall, "Sentimental Strength Detection for the Social Web", Journal of the American Society for Information Science and Technology. 63:163-173, DIO: 10.1002/asi.21662, Januart 2012

[3] Bryan Nii Lartey Laryea, Chi-Hwan Choi, In-sun Jung, Kyung-Hee Lee and Wan-sup Cho, "Web Application for Sentiment Analysis Using Supervised Machine Learning", International Journal of Software Engineering and its Applications, vol. 9, no. 1, pp. 191-200, 2015.

[4] Pan, IEEE Transactions on image processing, Vol. 20, N0.3, March 2011.

[5] Liu, X., & Zhang, Y. . Sentiment Analysis of Amazon Reviews Using Natural Language Processing Techniques.In Proceedings of the 2017 6th International Conference on Software and Computer Applications (pp. 171-175). ACM,2017.

[6] Mukherjee, T., Choi, I., Banerjee, U. Genetic analysis of fibroblast growth factor signaling in the Drosophila eye,2012.

[7] Liu, X., & Zhang, Y. . Sentiment Analysis of Amazon Reviews Using Natural Language Processing Techniques. In Proceedings of the 2017 6th International Conference on Software and Computer Applications (pp. 171-175). ACM, 2017.

[8] Devika, M.D., Sunitha, Cª., Ganesh, A.: Sentiment analysis: a comparative study on different approaches. Proc. Comput. Sci. 87, 44–49 (2016).

[9] Y. X. Liu, S. Goto, and T. Ikenaga, “A contour-based robust algorithm for text detection in color images,” IEICE Trans. Inf. Syst., vol. E89-D, no. 3, pp. 1221–1230, 2006.

[10] Neha S. Joshi, Suhasini A. Itkat, “ A Survey on Feature Level Sentiment Analysis” (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5422-5425.

[11] He Y., “Incorporating sentiment prior knowledge for weakly supervised sentiment analysis “, ACM Transactions on Asian Language Information Processing, Vol. 11(2).

[12] Bhadane, C., Dalal, H., Doshi, H.: Sentiment analysis: measuring opinions. Proc. Comput. Sci. 45, 808–814 (2015).

[13] M. Annet and G. Kondrak. A comparison of sentiment analysis techniques: Polarizing movie blogs.Advances in AI: 21st conference of the cancdian society for computational studies of intelligence, Canadian AI 2008, Windsor, Canada, May 2008, Proceedings.

[14] Hutto, C. J., & Gilbert, VADER: A parsimonious rule-based model for sentiment analysis of social media text. Eighth International Conference on Weblogs and Social Media (ICWSM-14),2014.

[15] Kang, H. W., Lim, J. H., & Kim, S. K. Comparative study of deep learning models for sentiment analysis of Amazon reviews. Journal of Information Science Theory and Practice, 9(2), 27-38,2021.