IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

Sentiment Analysis using Machine Learning

Ranganatha K[1], Chaitanya Kini S[2], Krishnamurthy Pai[3], Nakul Shet[4], Sudhir Hegde[5]
Assistant Professor, Department of Information Science and Engineering[1]
Students, Department of Information Science and Engineering[2,3,4,5]
Canara Engineering College, Bantwal, Karnataka, India

Abstract: *With the rise of e-commerce, consumers are becoming accustomed to shopping online and leaving reviews of their experiences on retailer and review websites. These op-ed pieces provide future customers with useful information for making decisions, and they also help businesses improve their goods and/or services. Yet, when the number of reviews increases quickly, users are forced to deal with a serious information overload issue. Several opinion mining techniques, such as opinion summarization, opinion polling, and comparative analysis, have been suggested to address this issue. How to correctly estimate the sentiment direction of review sentences is the main challenge. Sentiment analysis is a quickly developing subfield of the study of Natural Language Processing (NLP).In recent years, it has drawn a lot of interest. Sentiment analysis is used to assess or investigate user comments in order to make judgements about their opinions. Sentiment analysis is a machine learning (ML) technique in which computers classify and investigate human attitudes, feelings, and viewpoints towards the things that are communicated through text, star ratings, thumbs up and down, and other textual or graphical representations. The data for this study came from online product reviews acquired from the model website we created. Adjectives and adverbs can convey the opposite feeling with the use of negative prefixes. Such words are located using a negative phrase identification method. By using evaluation measures, the performance is assessed. Finally, we provide insight. Finally, we also provide a preview of our upcoming sentiment analysis research.*

Keywords: Sentiment Analysis

I. INTRODUCTION

Sentiment analysis is a collection of methodologies, techniques, and tools for detecting and extracting irrational information from language, such as opinions and beliefs, in order to assist in figuring out how buyers feel about purchasing a specific product or problem. It requires developing a system to gather and evaluate customer reviews of products posted on various online shops. By eliminating the human work required to read every single product review on the website, this system offers to create a classifier to handle the aggregation of these mixed feelings and provide a representation of these comments in a manner that is simpler and easier to understand.

II. METHODOLOGY

Figure 1: Control flow diagram for the proposed system

2.1 Data Collection

The data for this study was gathered by scraping user reviews for a specific product from the Amazon website using web scraping techniques. We chose Amazon as a data source for our study because it is a popular online marketplace with a wide range of products and reviews. To make an end point request to the specific page, we used Node JS-based web scraping tools such as axios, and cheerio to automate the data collection process. We created scripts that navigated the Amazon website, accessed the product's reviews section, and extracted the review text, product image, basic specifications, and other relevant information. We also took care not to scrape any of the reviewers' personal

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

information and to ensure that the scraping did not harm the website or violate any of its terms and conditions.

After scraping the data, we ran some preliminary checks to ensure that it was of high quality and appropriate for our study. We removed any duplicates and reviews with missing or incomplete information. We also filtered out any reviews that were not in English because we were only interested in English reviews. Finally, we removed any reviews with low ratings because they were unlikely to provide meaningful insights into the product's sentiment.

We stored the scraped data in a MongoDB database for further analysis because it required a storage medium. Throughout the data collection and analysis process, we made sure to keep the data secure and confidential, and we followed all ethical guidelines and best practises in data collection and management.

2.2 Data Pre Processing

We performed several preprocessing steps on the data after collecting it from the Amazon website in order to prepare it for sentiment analysis. The first step was to eliminate duplicates and irrelevant data. Following that, we cleaned the text by removing special characters, numbers, and stop words. To reduce the dimensionality of the data, we also used tokenization and stemming. All of the functionality mentioned above to prepare data for sentiment analysis is included in the sentiment and natural package of node JS. We used several techniques to remove noise from the text data during text cleaning. Any special characters or numbers were removed because they were unlikely to contribute to the sentiment analysis. Stop words, which are common words that are unlikely to add value to the analysis, such as "the," "and," and "but," were also removed. Tokenization, which involves breaking the text down into individual words or tokens, was performed using the Natural Language Toolkit (NLTK) library.

We manually inspected a sample of the reviews after preprocessing to ensure that the preprocessing had not introduced any errors or removed any important information. We also checked the sentiment label distribution to ensure that we had a balanced dataset. Overall, the data preprocessing step was critical in ensuring the accuracy of the sentiment analysis and the quality of the data.

2.3 Sentiment Analysis

After preprocessing the user reviews, we used sentiment analysis to determine the sentiment polarity of each review. The sentiment analysis was carried out using the Natural Language Processing (NLP) package from NPM. This package analyses text data and classifies it as positive, negative, or neutral using a pre-trained model.

We used the sentiment analysis package to categorise each review in our dataset as positive, negative, or neutral based on the sentiment polarity detected by the package. The sentiment label was then saved in the database alongside the review text and other pertinent data.Toensuretheaccuracyofthesentimentanalysis,wemanuallyinspectedasampleofthe reviews to see if the sentiment label assigned by the sentiment analysis package matched our own assessment of the sentiment of the review.

2.4 Graphical Representation

We plotted the sentiment distribution of the reviews using Chart.js, a JavaScript library for creating interactive charts and graphs, after labelling each review with its sentiment polarity. We made a doughnut chart that displayed the number of positive, negative, and neutral reviews over time, allowing us to see the product's sentiment trends.

2.5 Evaluation

Following the display of the sentiment analysis results in charts, we selected a sample of reviews from the dataset and manually evaluated the sentiment of each review. This allows you to compare the sentiment analysis results' accuracy to you rown judgement. The precision and recall of the sentiment analysis algorithm were calculated. The proportion of accurately labelled positive reviews among all positive reviews is known as precision. Recall is the proportion of all favourable reviews that were accurately categorised as positive. Cross-validation can be used to test the sentiment analysis algorithm's accuracy on a new dataset. Divide the dataset into training and testing sets, and train the sentiment analysis algorithm using the training set.

We manually inspected a sample of the reviews after preprocessing to ensure that no errors or important information had been removed. We also examined the sentiment label distribution to ensure that the dataset was balanced. Overall, the data preprocessing step was critical in ensuring the accuracy and quality of the sentiment analysis

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

III. IMPLEMENTATION

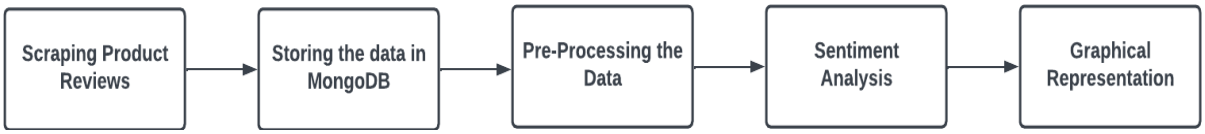

Figure 2: Data Flow diagram for the proposed system

We implemented our sentiment analysis project using Node.js, Cheerio, Axios, Sentiment and Natural NPM library and MongoDB. The implementation involved the following steps:

1. **Scraping product reviews**: We used the Axios library to make HTTP requests to the product page on Amazon's website and used Cheerio to parse the HTML and extract the product reviews. We collected the review text, date, and rating for each review and stored the data in a JavaScript object.
2. **Storing the data in MongoDB**: We used the MongoDB Node.js driver to connect to a MongoDB database and stored the review data in a collection. We created an index on the date field to make it easier to query the data later.
3. **Preprocessing the data:** We cleaned and preprocessed the review text by removing stop words, punctuation, and special characters. We also converted the text to lowercase and tokenized it into individual words.
4. **Sentiment analysis:** We used the sentiment library from NPM to label each review as positive, negative, or neutral based on the sentiment of the text. We stored the sentiment label as an additional field in the review data.
5. **Graphical Representation**: We used Chart.js library to create a bar chart that displays the sentiment distribution of the reviews over time. The chart is interactive allowing users to hover over each bar and see the exact number of reviews for each sentiment label

IV. RESULTS AND DISCUSSION

The proposed system for sentiment analysis is designed to provide a proper sentiment for each product based on many reviews considered in datasets. This Desktop application software will accept a url of the product and generate analysis of the product based on its product reviews on the e-commerce website. This analysis is done by running the product review through a classifier to judge the product whether its feedback is either positive, negative or neutral.. The proposed system is capable of providing a fully functional Desktop Application software to generate an analysis of product reviews, Proper validation to check if the URL entered is of a product listed on the e-commerce platform, Graphical User Interface for the user to interact with the app, To best identify the sentence as positive or negative, the most pertinent elements should be extracted. The analysis of the quality of the internet products is done using this positive and negative separation of remarks. To make an effective API to scrape reviews from the e commerce websites including the time, date, and stars.

V. CONCLUSION

This paper is proposed Sentiment analysis using machine learning which is developed to provide a proper sentiment towards a product based on many reviews on the product. In this study, it was discovered that opinion mining and sentiment analysis are crucial when choosing between various goods and services. Yet, it is crucial to take into account specific quality indicators while examining each review, such as helpfulness, usefulness, and utility. Sentiment classification or opinion mining may be used to any fresh applications that follow the rules of data mining. Although the algorithms and methods used for sentiment classification are rapidly improving and generating high-quality results, many questions in this field of research are still unanswered, making it difficult to identify fake reviews simply by reading them. Sometimes fake reviews are posted in place of real ones, while others are changed to make it impossible to determine their true motivations. Hence, identifying fake reviews is another vital area that need sophisticated data mining techniques

REFERENCES

[1]. Xing Fang ,Justin Zhan ,"Sentiment Analysis using product review data", Springer: Journal ofBigdata",2015,North Carolina A& T State university, Greensboro, NC, USA.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 2, March 2023
Impact Factor: 7.301

[2]. Subhabrata Mukherjee, Pushpak Bhattacharyya, "Feature Specific Sentiment Analysis forproductReviews",IET,2015,IIT Bombay.

[3]. Jian Jin and Ping Ji," Mining online product reviews to identify consumersFineGrinedConcerns",IET,2015,Hong Kong SAR, China.

[4]. Lada Banic,Anamihanovic,MarkoBrakus,"Using Big Data and Sentiment Analysis in Product Evaluation "MIPRO,2013,Croatia.

[5]. Haruna isah, Paul Trundle, DaneielNeagu, "Social Media Analysis for Product Safety and using Text Mining and SA",IET,2015,University of Bradford, UK.

[6]. p.c, Shilpa & Shereen, Rissa & Jacob, Susmi& Vinod, P.. (2021). Sentiment Analysis Using Deep Learning. 930-937. 10.1109/ICICV50876.2021.9388382.

[7]. Başarslan, Muhammet&Kayaalp, Fatih. (2021). Sentiment Analysis with Machine Learning Methods on Social Media. Advances in Distributed Computing and Artificial Intelligence Journal. 9. 5-15.10.14201/ADCAIJ202093515.

[8]. D. M. Thomas and S. Mathur, "Data Analysis by Web Scraping using Python," 2019 3rdInternationalconference on Electronics, Communication and Aerospace Technology (ICECA), 2019, pp. 450-454, doi:10.1109/ICECA.2019.8822022.

[9]. Pankaj, P. Pandey, Muskan and N. Soni, "Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), 2019, pp. 320-322, doi: 10.1109/COMITCon.2019.8862258.

[10]. Jagdale, R.S., Shirsat, V.S., Deshmukh, S.N. (2019). Sentiment Analysis on Product Reviews Using Machine Learning Techniques. In: Mallick, P., Balas, V., Bhoi, A., Zobaa, A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 768. Springer, Singapore.https://doi.org/10.1007/978-981-13-0617-4_61