International Journal of Recent Advances in Multidisciplinary Topics
Volume 2, Issue 6, June 2021
https://www.ijramt.com | ISSN (Online): 2582-7839

Sentiment Analysis Based Product Recommendation System

Charan Pote[1], Pranay Makode[2], Amarjeet Yadav[3], Kalyani Vasant Wadhai[4*], Divya Wanjari[5], Abhishek Pethe[6]

[1]*Professor, Department of Computer Technology, Priyadarshini Colleges of Engineering, Nagpur, India*
[2,3,4,5,6]*Student, Department of Computer Technology, Priyadarshini Colleges of Engineering, Nagpur, India*

***Abstract*: E-commerce sites are now a day having boom in selling and purchasing products, some of the most popular e-commerce sites are Amazon, flipkart etc where most of the customers visit to purchase products. Product reviews (genuine comments from the customers) is very important for customers, seller, businesses and manufacturers. Seller often want to know in time what consumers and the public think of their products and services. However, it is not really very feasible to manually read every post on the website and extract useful viewpoint information from it because there are so many comments about a single product on any e-commerce site. If you do it manually, there is too much data and it consumes your huge time. Sentiment analysis allows large-scale processing of data in an efficient and cost-effective manner to analyze the sentiments and conclude the result. In order to explore more about sentiment analysis, this paper tries to use the power of sentiment analysis to help the buyer and seller both. It helps buyers to see the honest reviews of the customers which already had purchased some product in the past and also to the seller to analyse the bunch of reviews which is being posted by the customers who purchased the product and based on this analysis they can improve their product for better selling.**

***Keywords*: sentiment analysis, amazon customer reviews, classification.**

1. Introduction

Amazon is one of the biggest online shopping site in the world. People often trust over the products and reviews of the product before buying the product on amazon itself. Also for seller it a beneficial site as a large number of peoples visit this site for shopping so seller is also getting a huge number of customers there, but the seller also has to improve its product quality as the customers want, to satisfy the customers. The proposed model helps the buyers to get the analysis of the reviews of the past customer of that product so that he/she can quickly take the decision whether the product is worth to by or not instead of wasting time in scrolling all the comments and reading it, it consumes time. Reviews sentiment analysis is the process of determining whether a piece of writing is positive, negative or neutral. Sentiment analysis helps data analysts within large enterprises to monitor brand and product reputation, and understand customer experiences.

1. Example: “I really like the new design of your website!” → Positive.
2. “I’m not sure if I like the new design” → Neutral.
3. “The new design is awful!” → Negative.

A product review is nothing but thehonest comments of the person who already used that product, our model is trying to collect all this review and trying to analyse it to help the buyer to take decision about buying the product. For example, if the user writes in his review, “the laptop is giving very high performance and I am satisfied with the product.”, then we can classify this comment as positive because customer is happy with the product. We aim to build a system that analyse the review’s sentiment in and come to the result whether to buy the product or not.

2. System Flow

1) Step 1: Web Scraping

In this step, data are fetched data from the e-commerce site here it is amazon. The data is particularly the reviews (comments) which is given by the customer who experience the performance of product.

2) Step 2: Review Collection

In this step, multiple reviews are collected from the web page and store it in a Test.csv file.

3) Step 3: Data Pre-Processing

In this step, the comments are captured from the Test.csv file and remove all the unwanted data which does not contribute in analyzing sentiments.

This includes: -

a) Removing all the handles like @,#
b) Removing punctuation symbols like?, %,&,*,+,=,-,*,/
c) Removing short words like are, is, am, are, were, was, will, shall
d) Next is tokenization which include splitting a phrase, sentence, paragraph, or an entire text document into smaller units, such as individual words or terms.

4) Step 4: Data Analysis

In this step, the pre-processed data are analyzed by comparing it with the data set in our model.

5) Step 5: Sentiments classification

This step calculates the final sentiment based on the

*Corresponding author: kalyaniwadhai240@gmail.com

prediction of sentiment analysis on each individual comments in the Test.csv file which is being generated after web-scraping.

6) Step 6: Conclude result to user

This step is simply displays messages to the user whether product is good to buy or not by the previous analyzation which is being done in the previous steps.

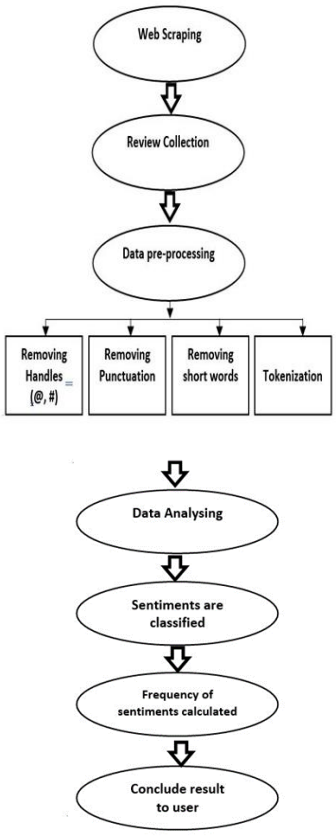

Fig. 1. Flow of execution

3. Proposed Algorithm

1) Web scraping (Used for comment collection of target product)

1. Load the amazon home page url. This is to search the product in which user is interested.
2. Locate the search bar and search button in home page. Enter the product name entered by user in the system's GUI in search bar of amazon home page and hit the search button by automation. This step redirect system to the next page having list of products which contain the target product as well as the related product.
3. Locate target product component in this random product list, and hit on that component for collecting comments of target product. After hitting enter it will redirect the system to next page and this is the main target page, where the system will get comments
4. Crawl the target product page url for the user entered product to extract all comments from it. Special care for required format of information must be taken, example tags have a special meaning to the browser i.e. break read or next line, we need to explicitly convert each tag to spaces or else the crawling result will be improper.
5. Store all the fetched comments in the "Test.csv" file.

2) Sentiment analysis on test. Csv file

6. Cleaning the crawled data. Removal of all special characters (such as: ":/.,'#$*^&-) must be done in order to retrieve best results. This also saves our review processing time. Put the crawled content into a Test.csv file.
7. Read the Test.csv file for processing, and for each comment in the csv file we perform the sentiment analyzation to find the frequency of the positive, negative and neutral emotions
   i. For each review we extract its sentiment from the comments using rule based extraction (using regular expressions).
   ii. Each sentiment extracted above is then sent to polarizer that return 1 if the sentiment is positive else -1 which means the sentiment is negative.

   A bar graph is plotted to finalize the result.
8. Final result is generated with the classification and sentiment of all the comments.
9. This result is then displayed in GUI.

4. System Working

The system working starts and ends on GUI, and the internal processing is based on the major 2 modules.

1. Input taken from GUI.
2. Internal processing.
3. Output displayed in GUI.

As shown in the figure, working of the system starts from taking the product name as input from the user, once the user throws the product name the system starts finding that product in amazon for fetching the comments of that comment. The system fetch comments from multiple comment pages of a single target product and further perform sentiment analysis on that comments to extract the resultant output by analysis the positive, neutral and negative sentiments in the system. Based on this sentiment analysed the system is going to display the user to final output in the form of message like. If the sentiments are

- Positive: You can buy this product.
- Neutral: Neutral sentiments.
- Negative: You can look for some other product

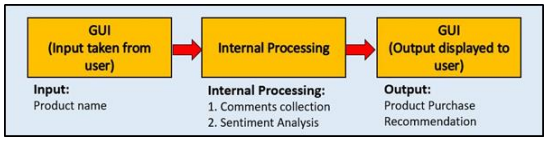

Fig. 2. System working

A. Complete Execution of system

1) Input taken from GUI

Interface of the system is built in tkinter, tkinter package ("Tk interface") is the standard Python interface to the Tk GUI toolkit. Both Tk and tkinter are available on most Unix platforms, as well as on Windows systems. System Interface is

having one input bar to take product name or either product url as input from the user, a show result button to start the processing.

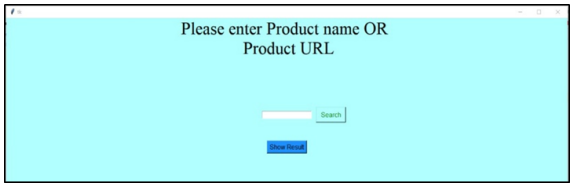

Fig. 3. Input taken from GUI

2) Internal processing.

Once the user enters the product name and hit on show result button in the interface the internal processing of the system starts. Internal processing, starts with web scraping part which performs the comment fetching task and next important task is to perform sentiment analysis on this collected set of comments to optimize the result.

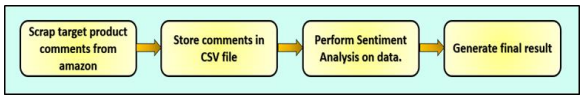

Fig. 4. Internal processing

B. Comment collection using web: scraping

System is using amazon (https://www.amazon.in/) site for searching and scraping comments of the user entered product. This is the most trusted site all across the world, so system can easily get the genuine and honest reviews about any product in this site, as most of the peoples purchase the products from this site and posted there review in the site about their experience with the product. For this web scraping part system is using certain python modules, they are as follows:

- Selenium
- bs4
- Requests
- Some user defined modules

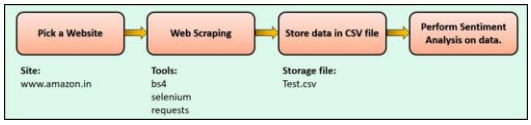

Fig. 5. Comment collection using web scraping

Selenium is used for web automation with the support of python, this system is using it to automate the searching of any product in amazon and reach up to the target product page from the amazon home page by entering the product name automatically in the search bar for which the user wants to find the sentiments analysis of comments and hitting the search button on the amazon home page and reaching to the target product page where the comments of product are present.

bs4 is used for extracting the tags and its data from the HTML code of web page also it is helpful to convert the HTML code in parse tree for better traversal of data.

C. Use Beautiful Soup to parse the HTML content

To parse HTML document and extract the div containers, we'll use a Python module called BeautifulSoup, the most common web scraping module for Python. Requests is a Python library used to easily make HTTP requests. Generally, Requests has two main use cases, making requests to an API and getting raw HTML content from websites (i.e., scraping). Some other user defined modules are

Scraper_Initializer is having Initialize_ fetcher (product_name) function which is taking product name as input and performing the automation on amazon site for fetching the url of the page having the target product and related product in the page as shown in fig 6.a. And further finding the url or the target product page as the page shown in fig 6.b.

Get_Next_Comment_Pages_Links is having Get_Next_Comment_Pages_Links (url, no_of_pages) function which is taking the url target page as shown in the fig 6.b and no of pages and find link of all the next page which we get after clicking on "Next Page" button in bottom of every comment page in site and returns the list of all Next page link to fetch comment from it.

Redirect_to_product_page is having Redirect_ to_ product_ page (url, product_name) function which is used to find the target page url of particular product from the list of multiple products.

Figure 6: Enter product name in search bar in Home Page

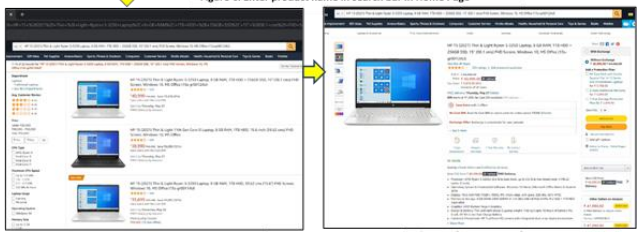

Fig. 6(a). Related product list page

Fig. 6(b). Target product page

Get_Next_Comment_Pages_Links is having Get_Next_Comment_Pages_Links(url, no_of_pages) function which is used to fetch the url of all the comments pages which usually get after clicking the "Next Page" button on the site. fetch_href is having fetch_href(a_tag) function which is used to find the href from any anchor tag.

Ex: -

This anchor tag is having one href in it which is fetched by fetch_href () function.

"https://www.amazon.in/HP-3-3250-Laptop-Windows-15s-

gr0011AU/dp/B08T6SJ215/ref=sr_1_1?dchild=1&keywords=HP+15+%282021%29+Thin+%26+Light+Ryzen+33250+Laptop%2C+8+GB+RAM%2C+1TB+HDD%2C+15inch+FHD+Screen%2C+Windows+10%2C+MS+Office+%2815s-gr0011AU%29&qid=1621000957&sr=8-1"

Web Scraping starts from Initialize_fetcher () function which is going to load the chrome driver into the script and web driver from selenium to use the chrome into the system for accessing the amazon site. After this system is going to requests for https://www.amazon.in to the server by using driver. Get () function after this by using driver we can parse the HTML tree of the website to fetch the data of our interest from the site code. After fetching the comments from the amazon site from the target product it is stored in "Test.csv" file as shown in fig 7

D. Sentiment Analysis

1) Converting text to lowercase

The very first step in sentiments analysis is to convert all the text into lowercase. The problem that can be raised if you don't convert your text completely into lowercase or uppercase. E.g: Apple with just an uppercase A, APPLE all in uppercase letters, or ApPle with a mixture of uppercase will all be counted separately. System want to change the text so that all the three versions of apple will be counted as same word.

2) Removing Punctuation

Punctuation can also cause problems. The basic approach is to deal with this is to remove everything that isn't a standard number or letter. However, if our text consists of punctuation like #, @, $ that doesn't spell any emotion so for better pre-processing of our text system should remove it from the text.

3) Tokenization

Tokenisation is the process of breaking up a given text into units called tokens. Tokens can be individual words, phrases or even whole sentences. In the process of tokenization, some characters like punctuation marks may be discarded. As the sentiments of a product can be evaluated using the words only so it is necessary to break the sentence and create tokens.

Fig. 7. Test.csv file

4) Removing Stop words

Stop words are the words in any language which does not add much meaning to a sentence. They can safely be ignored without sacrificing the meaning of the sentence. E.g: The words like I, we, they, for, thus etc. That doesn't spell any kind of meaning or emotion. So, for efficient pre-processing system is removing stop words from the input text.

5) Lemmatization

Lemmatization is the process of converting a word to its base form. The difference between stemming and lemmatization is, lemmatization considers the context and converts the word to its meaningful base form, whereas stemming just removes the last few characters, often leading to incorrect meanings and spelling errors.

6) NLP Emotion Algorithms

There are some of the words for e.g: love, like, affected, awesome which indicate a common emotion 'happy'. System can take this situation from a customer point of view if a customer likes a certain product then he or she will obviously give their reaction towards the product from the above words. So, what the emotion algorithms is doing it is looking for words that spells similar meaning. So, that system can store a common simple meaning for that words and after that the algorithm is counting how many times the words occur in the sentence.

Fig. 8. Emotion list

7) Sentiment Decider Code

Here, system have defined a function that is counting the intensity of the text. Whether the code belongs to positive, negative, neutral sentiments and depending upon the polarity of the text it depicts whether you should buy the product or not.

8) Output displayed in GUI.

After internal processing we got the result to show user, whether he/she can purchase that product or not. In the output, system displays one bar graph which shows the frequency of the words which contributes for sentiment analysis from comments. Also, it displays the recommendation message based on positive, negative and neutral sentiment.

Fig. 9. Output displayed in GUI

As in the above figure positive sentiment is analyze for product entered in the search bar so the message is displayed to user is "You can buy this product" as shown in fig 9.

5. Future Scope

The future of sentiment analysis based product

recommendation system is going to continue to dig deeper, far past the surface of the number of likes, sentiment of comments and aim to reach, and truly understand, the significance of social media interactions and what they tell us about the consumers opinion behind the screens. Efficiency of this system can be improved by replacing the rule based sentiment analysis process by Machine learning algorithms and manual web scraping of comments can be replaced by APIs provided by amazon, so that the execution time is decreased, and efficiency is increased.

6. Result

The main objective of this paper is to ensure fair results of sentiments, also don't want users to spend a lot of time reading through long textual comments in the reviews, and hence we summarize our result in the form of charts (Statistical Graphs). Data visualization is an important technology in the coming future, as data is increasing in size and complexity. Hence tis paper summarizes the results as bar charts that help users to view and directly understand the sentiment extracted also it displays the result in the form of a message. This paper is classified the reviews and doing a sentiment analysis on it.

7. Conclusion

The system is accurate enough for the test case of products reviews on amazon. For sentiment analysis, we have designed our own methodology that integrates existing sentiment analysis approaches. Classification of reviews along with sentimental analysis increased the accuracy of the system which in turn provides accurate reviews to the user.

References

[1] Aashutosh B, Ankit Patel, Harsh Chhedaet al, Amazon Review Classification and Sentiment Analysis/ (IJCSIT) *International Journal of Computer Science and Information Technologies*, vol. 6, no. 6, 2015.

[2] Emil Person ,Evaluating tools and techniques for web scraping, Degree Project In Computer Science and Engineering, Second Cycle, 30 Credits Stockholm, Sweden 2019

[3] D. Mali, M. Abhyankar, Sentiment Analysis Of Product Reviews For E-Commerce Recommendation ,*International Journal of Management and Applied Science*, vol. 2, no. 1, Jan.-2015.

[4] Abdullah Alsaeedi1, Mohammad Zubair Khan, "A Study on Sentiment Analysis Technique of Twitter Data", (IJACSA) *International Journal of Advanced Computer Science and Applications,* vol. 10, no. 2, 2019.

[5] Aliza Sarlan, Chayanit Nadam, Shuib Basri ,"Twitter Sentiment Analysis", 2014 International Conference on Information Technology and Multimedia (ICIMU), November, 18 – 20, 2014, Putrajaya, Malaysia

[6] Vishal A. Kharde, S.S. Sonawane ,"Sentiment Analysis of Twitter Data: A Survey of Techniques*", International Journal of Computer Applications*, vol. 139, no. 11, April 2016.

[7] Diksha Khurana1, Aditya Koli1, Kiran Khatter1,2 and Sukhdev Singh "Natural Language Processing: State of The Art, Current Trends and Challenges" ,Department of Computer Science and Engineering Manav Rachna International University, Faridabad-121004, India Accendere Knowledge Management Services Pvt. Ltd., India.

[8] Raja Selvarajan and Asif Ekbal, "IIT Patna: Supervised Approach for Sentiment Analysis in Twitter", Department of Computer Science and Engineering Indian Institute of Technology Patna, India.

[9] Prof. Alpa Reshamwala, Prajakta Pawar, Prof. Dhirendra Mishra, "Review On Natural Language Processing" IRACST – Engineering Science and Technology: *An International Journal* (ESTIJ), vol. 3, no.1, February 2013.