

Sentiment Analysis on Product Reviews using Web Scrapping

(Kajal Sharma, Harsh Sharma)¹, Er. Ankit Khare²

1UG student of Department of Information Technology, Shri Ramswaroop Memorial College of Engineering and Management Lucknow, Uttar Pradesh, India

2Assistant Professor, Department of Information Technology, Shri Ramswaroop Memorial College of Engineering and Management Lucknow, Uttar Pradesh, India

Submitted: 10-05-2022

Revised: 19-05-2022

Accepted: 22-05-2022

ABSTRACT – Nowadays the rise in demand for e-commerce with people preferring online purchasing of products, even we've got seen the rise in purchasing products online in these pandemic situation. There's an unlimited amount of data being shared. The e-commerce websites are loaded with large volume of knowledge. Even social media has become the nice help for selling and buying goods, and an excellent deal in sharing of this information throughout the world. This has influenced consumer's habits everywhere the world. Thanks to vivid reviews provided by the shoppers, there's a feedback environment being developed in e-commerce websites for helping customers buy the correct products and guiding the businesses to boost the features of products suiting consumer's demand. The sole disadvantages for the purchasers is availability of this huge volume of data and its non-uniformness structure. The customers find it difficult to exactly find the review for particular features of a product that they intended to buy. Also there are lots of review that they find difficult to read all and make decision for the product. We propose a technique to find the exact opinion about product and further classifying it into positive and negative reviews using web scrapping based approach that help customers to make decision for buying that product or not.

Key Words: Sentiment Analysis, Opinion Mining, Web Scrapping, Product Reviews, Natural Processing Language etc.

I. INTRODUCTION

Recently we've witness the rise of shopping for products or goods online and within the recent years E-Commerce websites has explored everywhere within the world. Majority of population is preferring to buy products through these websites online. Therefore the great amount

of information within the variety of reviews is produced which helps prospective buyers to decide on the proper product. Furthermore these reviews containing opinion of other customers about the products which might be useful for the corporate to spot the areas which require to be enhanced within the products. Besides the reviews and ratings provided do little to assess the particular features of the products. Thanks to all the Constrains the customer is unable to form a call about the product.

Sentiment Analysis is also known as opinion mining can be used to extract customer reviews from different sources on the internet. This technique implements various algorithms to analyze the corpus of data and make sense out of it. It helps to identify the orientation of a sentences thereby recognizing the element of positivity or negativity in it. Opinion mining can be implemented through a web scrapping based approach. Opinion mining uses natural language processing to extract the subjective information from the data (in this case its customer reviews for the product).

1.1 PROBLEM STATEMENT

A web application that collects reviews from particular e-commerce website where customer want to buy a certain product and analyzes them. It would segregate the reviews into positive and negative reviews. The reviews will be helpful to the customers to take decision for buying the product based on the user's feedbacks. The application further provides the pros and cons of the individual feature of the product and also provide reports about sentiment analysis performed on the products. We further aim to create a recommendation system that recommends products to users according to the feature requirement of users.

1.2 MOTIVATION

Very useful for recognizing the opinion of any particular product. Regardless of the name, the real motivation of sentiment analysis is the same, to know a user or audience opinion on a target object by analyzing a vast amount of text from various sources. We are able to analyze text on different level of details, and the detail level depends on customer's opinion. Sentiment analysis is already using in various organizations as Reputation Management, Social –Media monitoring, Brand Monitoring, Market Research, Competitor Analysis as well as Product Analysis. As I see in my surroundings I came to know many people don't like to read all those reviews, they are long or in large volume. They are unable to make decisions better so that they can use our Sentiment analyzing by following only few steps to get an efficient analysis which they want.

II. LITERATURE SURVEY

- Sentimental Analysis on online product review [2017]: In this paper, Raheesa Safrin, E.A. Vimal used feature vector classification for analysis of sentiment. The data used in this study is online product review collected from sample website that we have created word such as an adjective, adverb are able to convey opposite sentiment with the help of negative prefix, negative phrase identification is used. [1]
- Sentimental Analysis of Product review [2014]: In this paper, Aarti Patil used advanced Naïve Bayesian Algorithm technique to find whether the reviews are positive or negative for analysis of sentiment. The sentiment analysis is used to extract, aggregate and analyse the opinion on product from discussion forum. [2]
- Comparative Analysis of Sentimental Orientation using SVM and Naïve Bayes Techniques (2016): In this paper, Shweta Rana used Naïve Bayes classifier in system. Sentiment classified as positive or negative sentiment using film user review. Algorithm like Naïve Bayes, Linear SVM and Synthetic words is used. [3]
- Extracting a sentiment from Review: A Lexicon Based Approach [2017]: In this paper, Sujata Sonavane used Lexicon Based Approach for extracting sentiment such as positive, negative or neutral. Extracting the useful content from the opinion sources becomes a challenging task. To extract sentiment from review SentiWordNet is used to assign the polarity of sentiment. The classification of

review document is predicted by sentimental score. [4]

- Sentiment Analysis and Prediction of Online Reviews with Empty Ratings [2018] : In this paper, Sasikal, L. Mary Immaculate Sheela used Naïve bayes algorithm and logistic regression, Multinomial and Bernoulli classifier for the positive and negative reviews. [5]

III. TECHNOLOGIES

- **Python:** Python is a widely used dynamic programming language with a clean syntax and an indentation structure easy to learn [6].
- **Streamlit:** Streamlit is an open-source Python library that makes it easy to create and share beautiful, custom web apps for machine learning and data science.
- **NLP:** NLP is natural language processing tool. It has a comprehensive toolkit with a good range of grammar checking tools. It is fast and reliable. It identifies the part of speech of the words in a sentence. It is flexible and extensible
- **BeautifulSoup:** BeautifulSoup is a third party Python library from Crummy designed for scraping.

IV. METHODOLOGIES

The dataset used for this project is the Amazon Reviews Database [7]. The reviews in the dataset are consists of the attributes such as: Reviewer ID, Product ID, Review Text, Rating and time of the review. The main source of data used is the product reviews from Amazon. The reviews for a few popular phones have been obtained by building a web crawler. The web crawler has been written in Python using a scraping library called BeautifulSoup. Along with the review text, some additional data related to the reviews such as reviewer name, review date, overall rating and comments were also obtained. The crawler is called periodically to get the most up-to-date reviews. Each review is generally treated as a sentence or a group of sentences. They are cleaned and stored in a CSV file.

Then, sentiment analysis is performed on the preprocessed reviews and overall sentiment score for each review is generated. Further for feature extraction, there are two cases:

1. **Single Feature:** If the review contains only a single feature, then the sentiment score of the review is assigned to the feature.
2. **Multiple Features:** Some reviews have multiple features contained in them. So the above procedure will not work in this case. Rules are

defined to extract multiple features and assign the correct sentiment score to those features.

The Bayesian Classification is a supervised statistical method for classification and contains practical learning algorithms. The posterior probability of a class can be computed using Naive Bayes model. This model works is suitable for a large data set. The use of the Bayes Theorem is to presume the chance of the inclined feature set matches to specified label. Bayes theorem provides a way of calculating the posterior probability, $P(L|F)$, from $P(L)$, $P(F)$, and $P(L|F)$. Naive Bayes classifier assumes that the effect of the value of a predictor (F) on a given class (L) is independent of the values of other predictors. [8]

V. ARCHITECTURE OF PROPOSED MODEL

Architecture design gives the real world view of the system.

Data Flow Diagram represents the flow of information or data in the system. The square boxes represent the entities, ovals represent the process and named arrows represent the direction of flow of information. **Fig-1** shows the data flow diagram of web scrapping Sentiment analysis system. It has two entities, WebCrawler and Classifier and three processes namely Web Scrapping, Preprocessing and Product Review.

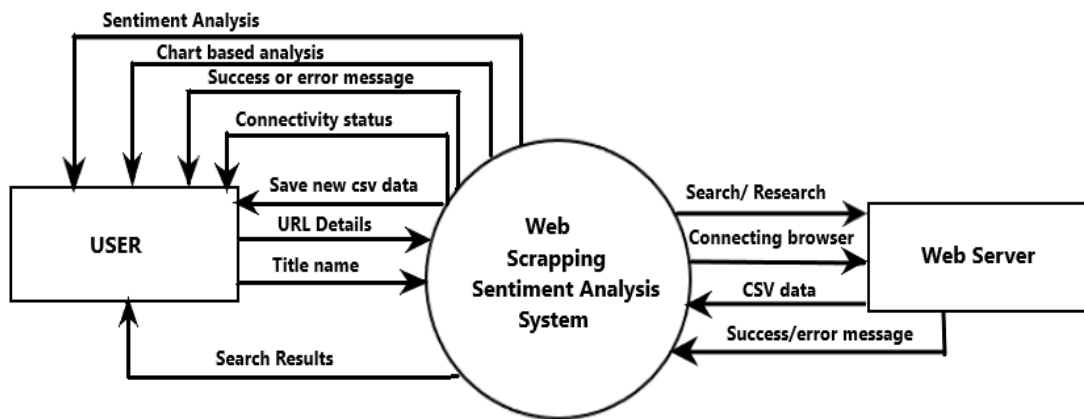


Fig – 1 Data Flow Diagram

Use Case Diagram is used show the interaction of user and system. The use cases and actors in use case diagram describe what the system does and how the actors use it, but not how the system operates internally. Fig-2 show actors as user and system and oval represents process.

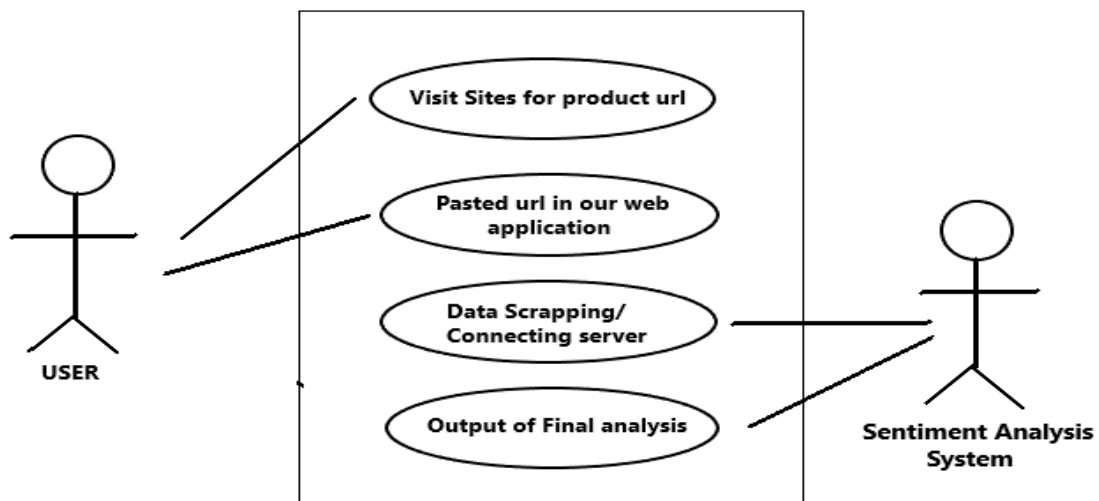


Fig-2 Use Case Diagram

VII. ADVANTAGES ANDDISADVANTAGES

- By using sentiment analysis, you gauge how customers feel about different areas of your product without having to read thousands of customer comments directly.
- If you've thousands of feedbacks per month, one person can't read all of these responses. By using sentiment analysis and automating this process you'll easily drill down into different customer segments of your business and procure a much better understanding of sentiment in these segments.
- The structure of websites change frequently - Scraped data is arranged according to the structure of the website. Sometimes you revisit the site and will find the ultimate layout changed.
- The downsides of Streamlit are its lack of design flexibility and control over your application layout.
- If your application and/or dataset is large, you'll likely run into speed issues due to the application flow - with the entire source code being re-run on every new change or interaction.

VIII. CONCLUSION

Sentimental analysis and opinion mining is a field of study that analyzes people's sentiment, attitude's or emotions towards certain entities. This project on sentimental analysis on product review using web scrapping tackles a fundamental problem of sentiment analysis, sentiment polarity categorization. Online product reviews from E-commerce website flip cart are selected as the input to this project. This will enable the system to analyze and then generate results (for reviews) which have been sorted and display in the form of charts. Possible future work is to improve the efficiency.

IX. ACKNOWLEDGEMENT

It is with great satisfaction and euphoria that we are submitting the paper on "Sentiment Analysis on Product Reviews". We are profoundly

indebted to our guide, Er. Akshay Tiwari and Er. Ankit Khare, Department of Information Technology, for innumerable acts of timely advice, encouragement and we sincerely express our gratitude. We also thank her for her constant encouragement and support extended throughout. Finally, yet importantly, we express our heartfelt thanks to our family & friends for their wishes and encouragement throughout our work.

REFERENCES

- [1]. Raheesa Safrin, K. R. Sharmila, T. S. Shri Subangi, E.A. Vimal "Sentimental Analysis on online product review", 2017
- [2]. Aarti A. Patil, "Sentimental Analysis of Product review," 2014
- [3]. "Comparative Analysis of Sentimental Orientation using SVM and Naïve Bayes Techniques" by ShwetaRana, 2016
- [4]. Sujata L. Sonavane, "Extracting a sentiment from Review: A Lexicon Based Approach," 2017.
- [5]. Sasikala , L.Mary Immaculate Sheela, Research Scholar, Department of Computer science, Mother Teresa Women's University, Kodaikanal, India. "Sentiment Analysis and Prediction of Online Reviews with Empty Ratings" ISSN 0973-4562 Volume 13, Number 14 (2018) pp. 11525-11531
- [6]. Python (programming language) - [https://en.wikipedia.org/wiki/Python_\(programming_language\)](https://en.wikipedia.org/wiki/Python_(programming_language))
- [7]. Inferring networks of substitutable and complementary products J. McAuley, R. Pandey, J. Leskovec Knowledge Discovery and Data Mining, 2015
- [8]. Evandro Costa, Rafael Ferreira, Patrick Brito, IglbertBittencourt, OlavoHolanda, Aydano Machado, TarsisMarinho "A framework for building web mining applications in the world of blogs: A case study in product sentiment analysis", Expert Systems with Applications (39),pp.4813-4834,2012.