



Sentiment Analysis Based Product Recommendation System: A Review

Prof. Charan Pote, Pranay Makode, Amarjeet Yadav, Kalyani Wadhai, Divya Wanjari, Abhishek Pethe
 Assistant professor at PCE, Student(PCE), Student(PCE), Student(PCE), Student(PCE), Student(PCE)
 Department of Computer Technology,
 Priyadarshini College 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 improv 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.

Example:

1. "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 the honest 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 final result whether to buy the product or not.

2. System Flow

Step 1: Web Scraping

In this step, data are fetched 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.

Step 2: Review Collection

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

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 analysing sentiments.

This includes: -

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

Step 4: Data Analysis

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

Step 5: Sentiments classification

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

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 analysis which is being done in the previous steps.

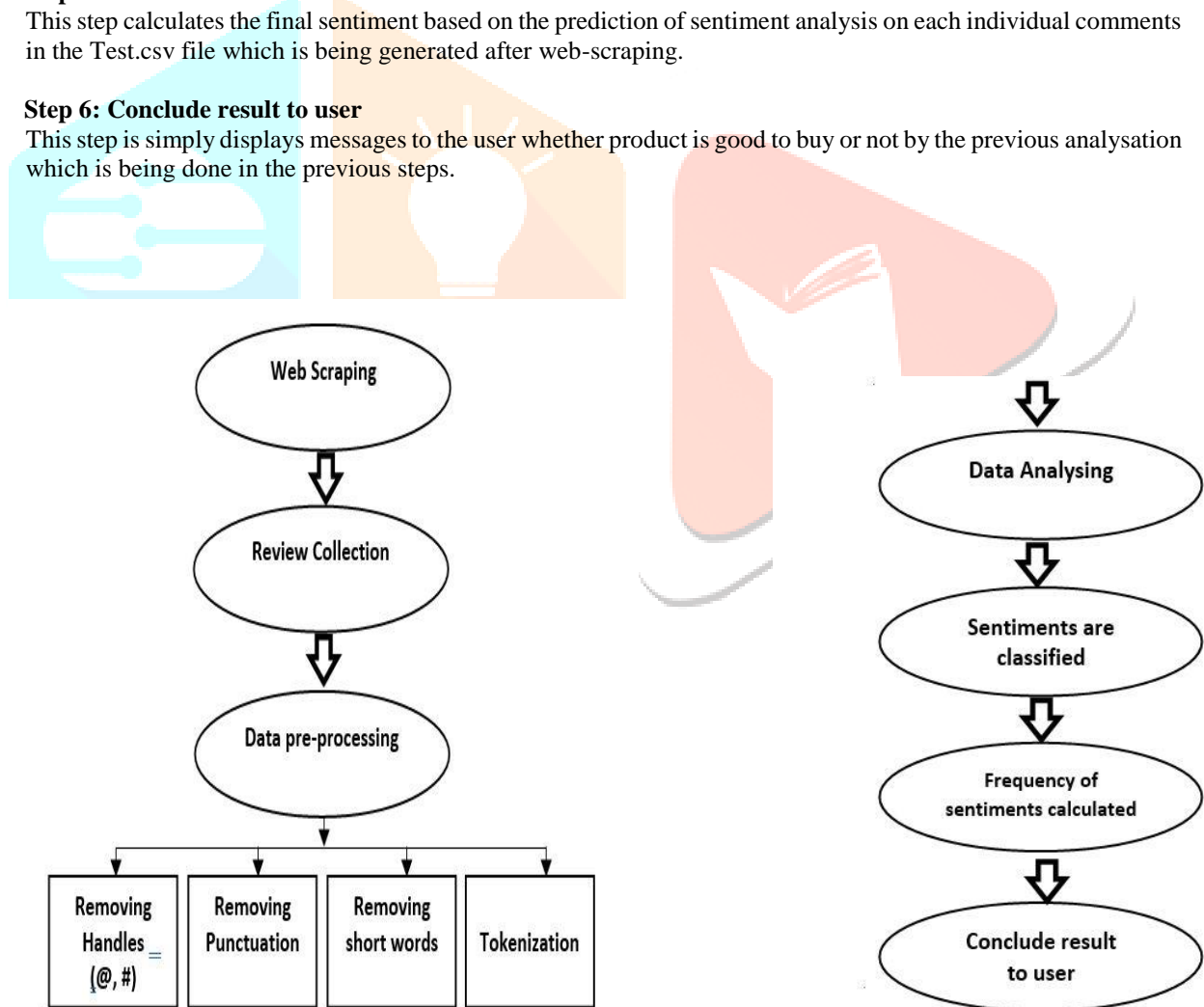


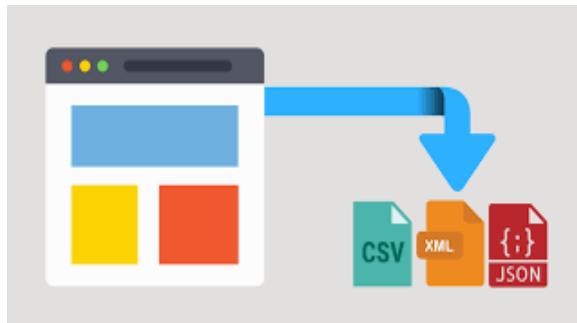
Fig 1. Flow of Execution

3. Modules

The sentiment analysis is divided into two modules

- 1) Web Scraping module.
- 2) Sentiment analysis module.

Module 1: Web Scraping



Web-scraping is the automated collection of information from webpages.

“The Web consists predominantly of unstructured text.

One of the central tasks in web scraping is to collect the relevant information for our research problem from heaps of textual data. Within the unstructured text we are often interested in systematic information – especially when we want to analyse the data using quantitative methods.

Systematic structures can be numbers or recurrent names like countries or addresses.

We usually proceed in three steps:

Step 1: Gather the unstructured text,

Step 2: Determine the recurring patterns behind the information looking for, and

Step 3: Apply these patterns to the unstructured text to extract the information.”

Example 1

- **Data:** food price data scraped from supermarket websites.
- **Use:** to produce timely measures of food-price inflation.

Example 2

- **Data:** jobs vacancy data scraped from jobs portals.
- **Use:** to produce timely jobs vacancy statistics and provide a richer source of labour market information.

Example 3

- **Data:** data related to second or holiday homes scraped from holiday lettings and room-sharing websites.
- **Use:** to help inform census and social survey design and estimation.

Example 4

- **Data:** text scraped from large numbers of business websites.
- **Use:** to produce research and statistics on the digital economy, for example, research into business classification systems.

Tools to be used for web scraping

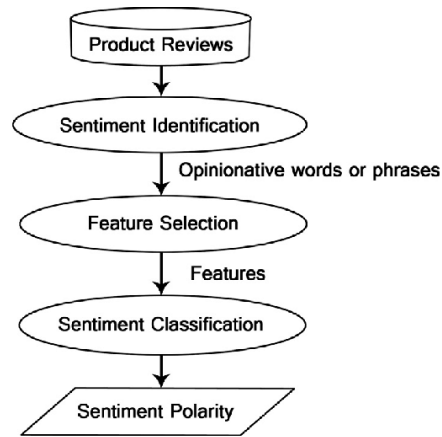
The web offers several tools that you can use. Most of them are free for limited use and offer paid plans for more advanced tasks.

R and Python are available, too. However, you should be trained in coding and comfortable using them.

In this paper, **python programming language** and some python module are used as web scraping tools like **selenium, BeautifulSoup, requests, lxml**.

Here, bunch of comments of the product are scrapped from amazon in which user is interested. This comment is being collected in the Test.csv file which is going to be the input for the sentiment analyser module.

Module 2: Sentiment Analysis



What is Sentiment Analysis?

Sentiment analysis is the process of detecting positive or negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, and understand customers.

Since customers express their thoughts and feelings more openly than ever before, sentiment analysis is becoming an essential tool to monitor and understand that sentiment. Automatically analysing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers' needs.

For example, using sentiment analysis to automatically analyse 4,000+ reviews about your product could help you discover if customers are happy about your pricing plans and customer service.

Maybe you want to gauge brand sentiment on social media, in real time and over time, so you can detect disgruntled customers immediately and respond as soon as possible.

The applications of sentiment analysis are endless

Types of Sentiment Analysis:

Sentiment analysis models focus on polarity (positive, negative, neutral) but also on feelings and emotions (angry, happy, sad, etc), urgency (urgent, not urgent) and even intentions (interested v. not interested). Depending on how you want to interpret customer feedback and queries, you can define and tailor your categories to meet your sentiment analysis needs. In the meantime, here are some of the most popular types of sentiment analysis:

Fine-grained Sentiment Analysis

If polarity precision is important to your business, you might consider expanding your polarity categories to include:

Very positive

Positive

Neutral

Negative

Very negative

This is usually referred to as fine-grained sentiment analysis, and could be used to interpret 5-star ratings in a review, for example:

Very Positive = 5 stars

Very Negative = 1 star

examples:

Emotion detection:

This type of sentiment analysis aims to detect emotions, like happiness, frustration, anger, sadness, and so on.

Aspect-based Sentiment Analysis:

Usually, when analysing sentiments of texts, let's say product reviews, you'll want to know which particular aspects or features people are mentioning in a positive, neutral, or negative way.

That's where aspect-based sentiment analysis can help

Why Is Sentiment Analysis Important?

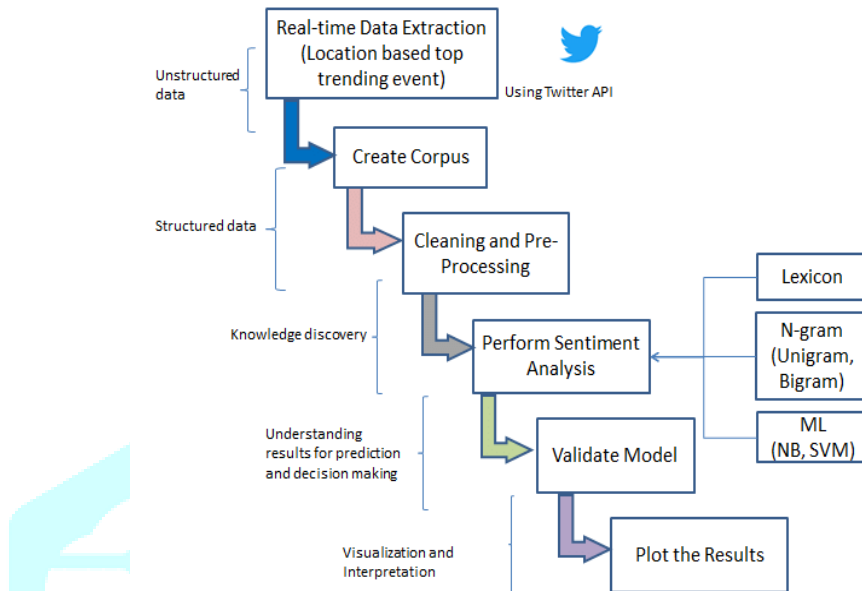
Sentiment analysis is extremely important because it helps businesses quickly understand the overall opinions of their customers.

By automatically sorting the sentiment behind reviews, social media conversations, and more, you can make faster and more accurate decisions.

It's estimated that 90% of the world's data is unstructured, in other words it's unorganized.

Huge volumes of unstructured business data are created every day: emails, support tickets, chats, social media conversations, surveys, articles, documents, etc).

But it's hard to analyse for sentiment in a timely and efficient manner.



How Does Sentiment Analysis Work?

There are different algorithms you can implement in sentiment analysis models, depending on how much data you need to analyze, and how accurate you need your model to be. We'll go over some of these in more detail, below.

Sentiment analysis algorithms fall into one of three buckets:

- Rule-based: these systems automatically perform sentiment analysis based on a set of manually crafted rules.
- Automatic: systems rely on machine learning techniques to learn from data.
- Hybrid systems combine both rule-based and automatic approaches.

Sentiment Analysis Challenges:

Sentiment analysis is one of the hardest tasks in natural language processing because even humans struggle to analyse sentiments accurately.

Data scientists are getting better at creating more accurate sentiment classifiers, but there's still a long way to go. Let's take a closer look at some of the main challenges of machine-based sentiment analysis:

1. Subjectivity and Tone
2. Context and Polarity
3. Irony and Sarcasm
4. Comparisons
5. Emojis
6. Defining Neutral
7. Human Annotator Accuracy

Sentiment Analysis Use Cases & Applications

The applications of sentiment analysis are endless and can be applied to any industry, from finance and retail to hospitality and technology. Below, some of the most popular ways are listed that sentiment analysis is being used in business:

- Social Media Monitoring
- Brand Monitoring
- Voice of customer (VoC)
- Customer Service
- Market Research

4. Algorithm

1. Crawl the amazon 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.
2. 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.
3. 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.
4. Final result is generated with the classification and sentiment of all the comments.
5. This result is then displayed in GUI.

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 final result in the form of a message. This paper is classified the reviews and doing a sentiment analysis on it.

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.

Reference:

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 (6) , 2015, 5107-5110
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, ISSN: 2394-7926 Volume-2, Issue-1, Jan.-2015,127
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 (0975 – 8887) Volume 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. Dharendra Mishra, “REVIEW ON NATURAL LANGUAGE PROCESSING” IRACST – Engineering Science and Technology: An International Journal (ESTIJ), ISSN: 2250-3498, Vol.3, No.1, February 2013.