

# Buy || !Buy: To Buy Or Not?

Aditi Khullar, Dharthi Nagaraj Kashyap, Parimal Deep Singh, Rajath Subramanyam

University of Illinois at Urbana-Champaign

## Abstract

In today's fast paced world, digesting and analysing the data quickly is the need of the hour. There is wealth of information on the World Wide Web and there are several choices for users. Particularly in online shopping and e-commerce, consumers are often flooded with choices. Very often consumers are not able to decide whether to buy products or not based on their specification or description. Reviews posted by other consumers can be used as a useful tool to make right decisions. On shopping sites like Amazon.com the reviews of certain products are huge and overwhelming to consumers. It is not possible to go through the reviews manually and make a judgement. There are several review summarizers that crawl through the reviews and either provide a summary of it or aggregate it in one place. However, there is no such tool that can aggregate the reviews and make an **intelligent decision** based on the reviews. In this paper we present such a tool that can help consumers make intelligent buying decisions based on customer reviews posted on e-commerce websites.

## 1. Introduction

Online shopping websites like Amazon [1] have several thousands of products sold on them. The same product are being sold by several different sellers [2]. Consumers are often spoilt for choices and are unable to decide based on the specification of the product or based on the description of the product. It is also highly time consuming if the user happens to be naïve in the category of product he or she is buying. A very powerful tool to make the right decision could be the feedback from consumers who have already got the product and have been using it for some considerable amount of time. Amazon provides opportunities to consumers to review products that they have bought or have used. There are several review

aggregators out there. But there are also equal number of hoax reviews often posted by sellers trying to bump up their products; or bots simply spamming the online shopping servers. So having mere review aggregators, putting all the reviews in one place isn't going to help. Also the consumers of today are accessing online shopping sites from mobile apps or mobile browsers and there isn't much space on the mobile display to even see the aggregation of reviews especially of the popular products. There is a need for a tool that saves users time i.e. a tool that not only aggregates reviews, but also filters false positives, takes care of discrepancies in reviews and finally makes an intelligent decision that consumers can trust. In this paper we present a product, *To Buy Or Not* (a.k.a. *Buy || !Buy* as a geek would

symbolize it) that aims to fill this gap. The product gives binary decision suggesting whether a user should buy the product based on the satisfaction expressed by others users who have reviewed the product on Amazon.com. Going forward, the tool can be expanded to aggregate reviews from other e-commerce websites like eBay [3], Best Buy [4].

### **1.1 Importance**

The tool will be useful to any user looking for a product and is not sure whether to buy the product based on the specification or the description of the product. This tool can help users who are undecided and help them make a decision based on reviews posted by other users who have used the product. This can also be useful to users who are very naïve in the category of the product and can save them a lot of time as opposed to going through the entire set of reviews one by one. Mobile users can also benefit from this tool too. Due to the limited display area it is hard to go through the entire set of reviews. In all these cases, these users can use our tool which can quickly perform an accurate sentiment analysis of all the reviews and in turn give an overall sense of satisfaction from the users who are already using the product and provide a binary decision on whether to buy the product or not. In case there is insufficient data to make a decision from like in case of no reviews, our tool prompts the user to “try their luck”.

### **1.2 Relevance**

Inspired by the common problem of digesting data, we were interested in making a tool mostly used by naïve users who do not want to understand or read up

too much before buying a product. There are too many products to read about, too many features and too many reviews. We want to aggregate this to a simple overall rating. There is also the question of a user's mood (anger or extreme happiness) while reviewing the product that they have just bought. This might affect the overall rating scale. We wanted to address this problem by doing sentiment analysis on the reviews. A straightforward sentiment analysis may not always say the truth. Since sentiment analysis is a hard problem [5, 6] we also take care of discrepancies in the sentiment of the review and check if it actually expresses that level of satisfaction and filter out false positives before making the final decision. For example, there might discrepancies in the rating given by a user and final outcome of the sentiment analysis run on his / her review.

### **1.3 Novelty of the idea**

In our opinion, there are not many popular generic review aggregators out there. Even if they exist they are limited to a certain category of products. Our product is beyond a review aggregator. It goes one step ahead, performs sentiment analysis on the set of reviews, filters false positives by taking care of discrepancies and provides an intelligent final decision.

Review summarizers exists but not all reviews on same products carry same sentiments. Two people voting a product with four stars may not imply the same level of satisfaction. Our product takes care of these discrepancies. It searches for relevant reviews, normalizes them, filters out false positives and false negatives, extracts relevant reviews, removes outliers

and redundant reviews and finally presents an intelligent binary decision to the user.

## 2. Implementation

To buy not to buy crawls and collects the reviews for the product that the user wants to buy. Once the user puts in the product she is interested in every review of that product is sent for sentiments Analysis. We built our own custom crawler in python to do this task. The task of sentiment analysis on the reviews is performed by AlchemyAPI[7].

### 2.1 Sentiment Analysis using Alchemy API

Sentiment is the attitude, opinion or feeling toward something, such as a person, organization, product or location. AlchemyAPI's sentiment analysis API provides easy-to-use mechanisms to identify the positive or negative sentiment within any document or webpage and in our case we use it for reviews of the products.

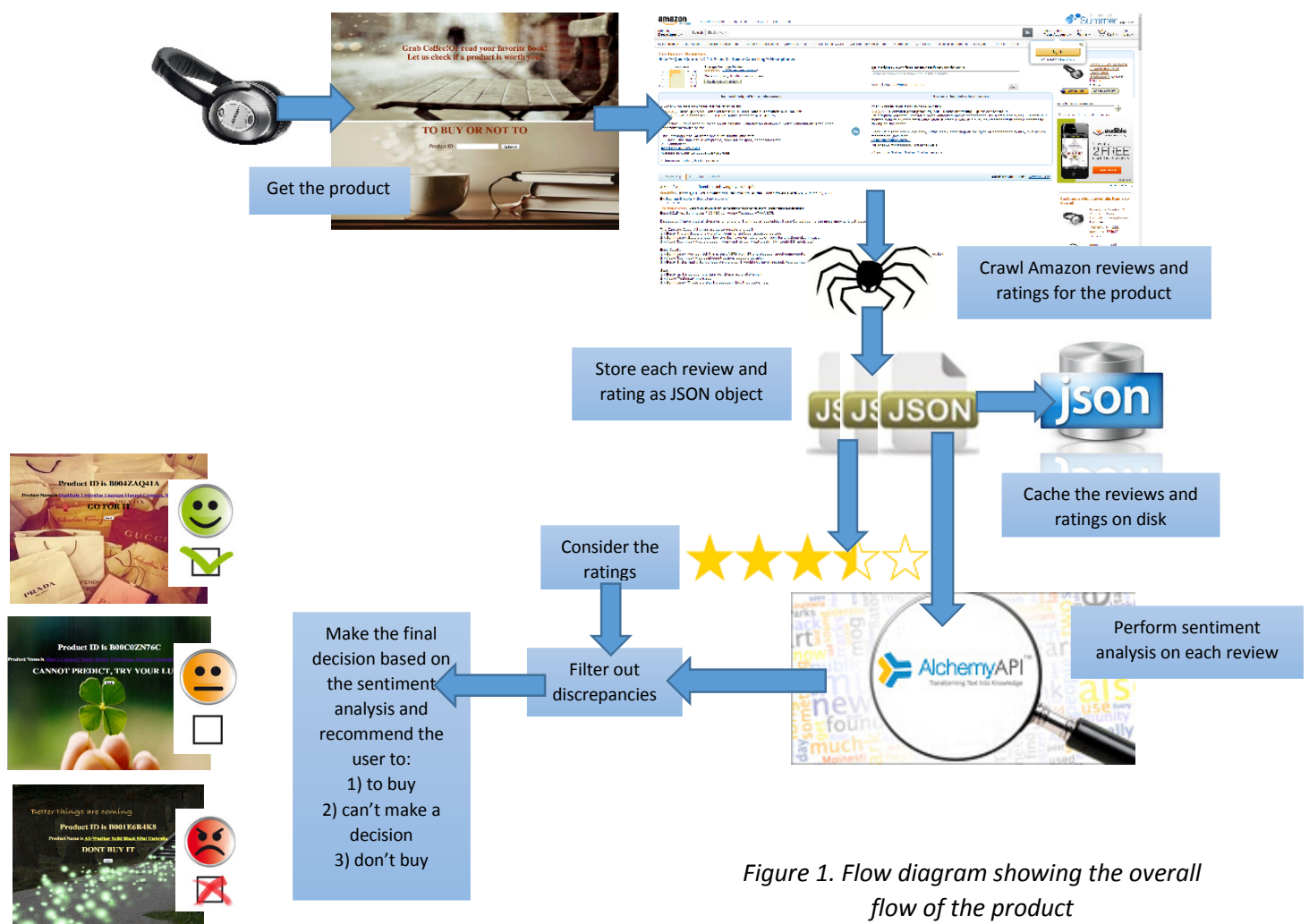


Figure 1. Flow diagram showing the overall flow of the product

The sentiment analysis API is capable of computing document-level sentiment, sentiment for a user-specified target, entity-level sentiment, quotation-level sentiment, and directional-sentiment and keyword-level sentiment. These multiple modes of sentiment analysis provide for a variety of use cases ranging from social media monitoring to trend analysis.

AlchemyAPI's sentiment analysis algorithm looks for words that carry a positive or negative connotation then figures out which person, place or thing they are referring to. It also understands negations (i.e. "this car is good" vs. "this car is not good") and modifiers (i.e. "this car is good" vs. "this car is really good"). The sentiment analysis API works on documents large and small, including news articles, blog posts, product reviews, comments and Tweets.

## 2.2 Review sentiments matched with review rating

Once we have the sentiment scores for each of the reviews, our software also fetches the review ratings (i.e. the count of stars which each reviewer gives to the product). The idea is to make sure that the sentiment analysed by the Alchemy API and the user are on the same lines. This also helps in putting another level of filtering and removes and false positive reviews in the dataset. For every review following steps are done:

1. Get the sentiment score from the Alchemy API
2. If the score is positive, we make sure that the rating of the product (stars of the product) is above at least 2.5

3. If the score is negative, we make sure that the rating of the product is below 2.5

4. The reviews which satisfy the criteria in 2 and 3 are the ones which are kept for further analysis

## 3. Evaluation

We ran some experiments to evaluate our product. One of the key attribute of a product is the how quickly it can provide a decision to the end user. In the first experiment, we ran the tool against a set of products with varied number of reviews and plotted the response time. Table 1 shows the products that were used for the analysis of the tool and the respective response time along with the final decision prompted by our tool.

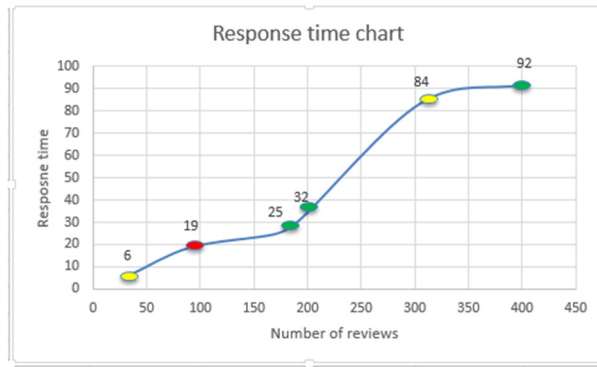
Product ID	Product name	# of reviews	Response time (seconds)	Final Decision
B008FQMW3E	Beats pill	402	92	Go for it!
B003M2YDE2	Metallic Monitor Stand	306	84	Try your Luck!
B00EHZYWGM	Amazon Basic portable speakers	194	32	Go for it!
B0072VOWH2	Babolat Tennis Racquets	170	25	Go for it!
B006107300	Lamb of god CD	91	19	Don't go for it!
B00FORE018	ASUS 15" Laptop	32	6	Try your Luck!

Table 1. Data points for the analysis.

Figure 2 shows the number of reviews against the response time. The data points are marked in different colours based on the decision suggested by the tool.

## 4. Discussion and Future Work

A great number of open source review aggregators exist online today. *To buy or not* is not just a review aggregator. It goes



*Figure 2. Chart showing the response times*

a step further. Not only does it summarize the reviews but also provides a final buy or do not buy decision to the user. Though we have built a comprehensive and unique tool that solves a major problem faced by customers who shop online these days, we would like to highlight some of the future enhancements that can be made.

To buy or not to presently take the product's ASIN (Amazon's Standard Identification Number) and does required analysis on the reviews. Working with ASIN adds an ambiguity check at the back end and the software knows exactly what the user wants to buy and fetches the right reviews. But at the same time fetching ASIN number for a product adds a level of complexity for the user. Hence, enhancing the software to take the product names and description instead of just the ASIN will be a great relief to the user of this product.

Apart from that, we also foresee this product fetching reviews not just from Amazon.com but also from the other ecommerce websites such as ebay.com and bestbuy.com. Making this product as a browser plugin where it can recognize

which website the user is browsing, and fetch corresponding results will be a great enhancement and add to the utility of the product.

Lastly, adding a feedback engine which takes the results and performance statistics from the past user behaviour and use that to improve the future recommendations can help in increasing the user satisfaction.

## 4. References

- [1] Amazon. [www.amazon.com](http://www.amazon.com).
- [2] Sell on Amazon. <http://www.amazon.com/gp/seller-account/mm-product-page.html>.
- [3] ebay. [www.ebay.com](http://www.ebay.com).
- [4] Best Buy. [www.bestbuy.com](http://www.bestbuy.com).
- [5] Why is sentiment analysis hard? <http://idibon.com/why-is-sentiment-analysis-hard/>.
- [6] R. Varghese, M. Jayasree. A survey on sentiment analysis and opinion mining.
- [7] AlchemyAPI. <http://www.alchemyapi.com/>.
- [8] C.C. Agarwal, C. Zhai. A survey on opinion mining and sentiment analysis. [http://link.springer.com/chapter/10.1007%2F978-1-4614-3223-4\\_13](http://link.springer.com/chapter/10.1007%2F978-1-4614-3223-4_13).
- [9] Esuli, A. and Sebastiani, F. Determining term subjectivity and term orientation for opinion mining. In *Proceedings of Conf. of the European Chapter of the Association for Computational Linguistics* (2006).
- [10] Ding, X., Liu, B. and Yu, P.S. A holistic lexicon-based approach to opinion mining. In *Proceedings of the Conference on Web Search and Web Data Mining* (2008)



## 5. Appendix

### 5.1. Screenshots of the tool

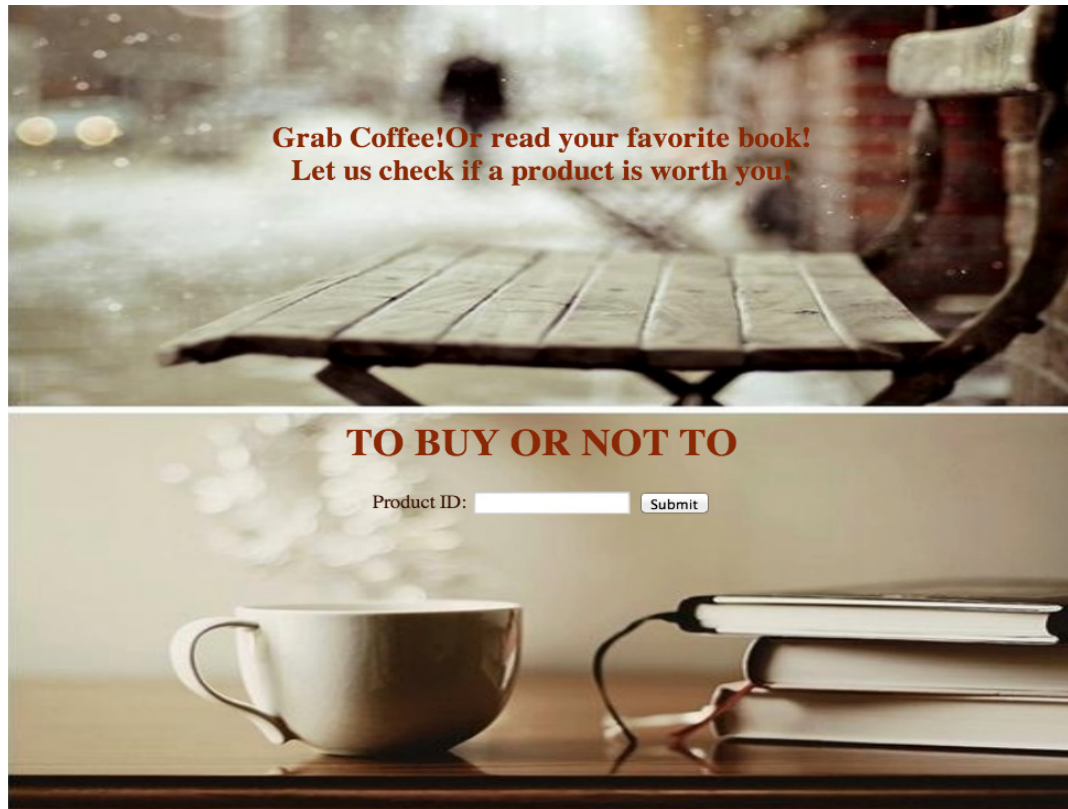


Figure 3. Main landing page of the tool



Figure 4. Positive results page ("Go for it!" 😊)



Figure 5. Negative results page ('Don't Buy it!' 😊)

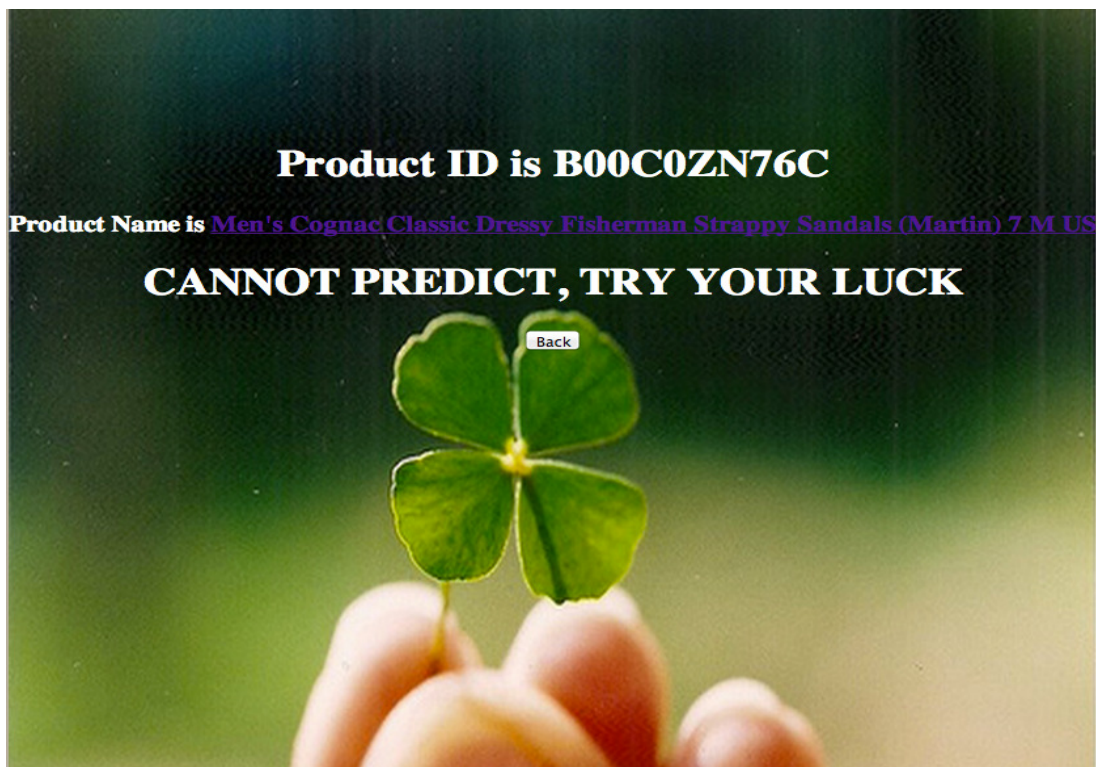


Figure 6. Neutral Results page ("Try your Luck!" 😊)

## **5.2. Work Division**

### **Implementation**

- **Amazon Crawler:** Rajath Subramanyam, Parimal Deep Singh
- **Alchemy API integration:** Parimal Deep Singh, Aditi Khullar
- **User Interface:** Dharti Nagaraj Kashyap, Parimal Deep Singh
- **Decision Making engine:** Aditi Khullar, Rajath Subramanyam, Dharti Nagaraj Kashyap

### **Project Presentation**

Dharti Nagaraj Kashyap, Parimal Deep Singh

### **Project report**

Aditi Khullar, Rajath Subramanyam