

LISTING2SPEAK: A DATA-DRIVEN ANALYTICAL SYSTEM TO EVALUATE THE E-COMMERCE PRODUCT LISTING USING ARTIFICIAL INTELLIGENCE AND BIG DATA ANALYSIS

Zihan Xu¹ and Yu Sun²

¹Crean Lutheran High School, 12500 Sand Canyon Ave, Irvine, CA 92618

²California State Polytechnic University, Pomona, CA, 91768,
Irvine, CA 92620

ABSTRACT

In e-commerce, product pages are important to the success of a website [4]. The ultimate goal of an e-commerce site is sales and it takes a great product page to achieve that. However, today, e-commerce entrepreneurs are confused about how to make their product page more attractive [5]. This paper designs a data-driven analytical system to analyze the relationship between different web page features with sales, in order to give users feedback on how to improve their product web pages [6].

KEYWORDS

Data Analysis, E-commerce, Artificial Intelligence.

1. INTRODUCTION

As the E-commerce system develops rapidly and becomes mature, e-commerce platforms have become an inseparable part of people's lives [7]. There is always an interesting phenomenon that two Amazon stores sell the exact same thing, but one sells much more than the other. In fact, for e-commerce websites, besides advertisement, the design of product pages is extremely important to the user's experience, and can even affect their purchasing decisions. First of all, the product details page is a product display area that can further stimulate customers' desire to buy. The biggest difference between online shopping and brick-and-mortar shopping is that customers cannot really touch the product, and can only rely on the product pictures displayed by the seller and the reviews of buyers who have already purchased to guess the quality of the product, and then decide whether to buy it or not [8]. In order to let products make better sales, the product details page of each product is our top priority. As a result, creating an E-commerce product page analysis project is necessary to help E-commerce sellers better improve their product web page. It brings users convenience and supports them to get better sales.

Some of amazon product listing optimization techniques and systems that have been proposed to identify the various factors that influence the success of your conversion and product listing sales, which allows the user's product to have a better ranking in its specific product category .These already existing tools claim to provide users with some of the best possible product

choices and ensure that their products are highly relevant when customers search for them by reading and analyzing all the data available on the market. However, these proposals assume that getting a higher ranking when customers search on Amazon is an absolute determinant of their sales, which is rarely the case in practice. Their implementations are also limited in scale because their system requires them to fill in product descriptions, keywords, etc. and upload images and videos on their own which will waste users a lot of time. Other techniques, such as “amify”. They offer an expert team to optimize customers’ Amazon Listing. Because people’s preference may change over time, the method of using people to optimize product listing often results in inaccuracy.

In this paper, we follow the same line of research by ... Our goal is to create an E-Commerce Product Listing reviewer based on a data analytical System [9]. Our method is inspired by one of machine learning ’ s functions: it can use algorithms to parse data, learn from it, and make decisions or predictions about something in the world which are compliant with our website’s goal. There are some good features both in our website page and inside our website’s backend. First, we divide different kinds of products into different categories, which can produce more accurate forecasts. Secondly, we use the data collected from Amazon to do Sklearn in order to analyze and find the relationship between aspects of product listings and sales [10]. Sklearn provides us the ability to transform these huge amounts of data to predictions and analysis results. Thirdly, we gathered a huge amount of data by using selenium to do web scraping. Although our computers do not allow us to collect all the sales data, top 100 products in each Amazon category can still enable us to make precise analyses.

The following of this paper is organized as follows: Section 2 details the challenges we encountered while writing the backend of the program and during our experiments; Section 3 highlights the details of our solutions to the challenges mentioned in Section 2 and the main METHOD used in our project; Section 4 presents the details related to the evaluation of our results by the experiments we performed, followed by the related work of our project in Section 5. Finally, Section 6 gives the conclusion and points out the existing limitations of this project and the aspects and work that can be improved in the future.

2. CHALLENGES

In order to build the project, a few challenges have been identified as follows.

2.1. Finding All the Features that can be Included in Our Website Page

The first challenge is finding all the features that can be included in our website page. In an Amazon product page, there is plenty of information such as description, pictures, videos and ratings [11]. Definitely I can not include all these features in my evaluation and feedback website because this would take our advice out of focus and some of the information can not convert to numbers in order to find a data relationship. To solve this challenge, my professor and I brainstormed on what users really want to make changes on their products. First, we eliminated ratings because ratings are something that cannot be changed by the design of the product page, as customers will only rate the product once they get the actual product, so the rating depends more on the quality of the product itself. Then we decided to include the number of keywords, description words, pictures and videos in our website which allows us to build AI models to analyze data relationships.

2.2. Difficulty Collecting Data on Amazon Sales

Amazon's whole product sales data is so huge that we can't collect all the sales data and do data analysis and future prediction with just the arithmetic power of our computers. We need a method that does not need to collect so much data but can provide a database for my ai model. We then found that collecting the top 100 lists for each product type on Amazon would provide both an adequate data set and not require excessive computer computing power.

3. SOLUTION

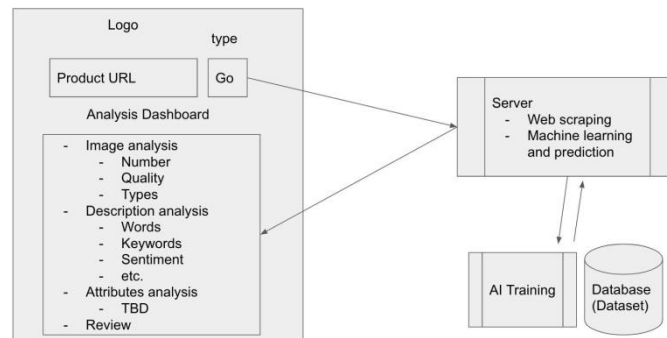


Figure 1. Overview of the solution

Listing2Speak is a Data-Driven Analytical System aimed to Evaluate the E-Commerce Product Listing using Artificial Intelligence and Big Data Analysis [12]. Listing2Speak is a powerful system for users to consider the way up their products' sales. It is able to analyze the input data and give users reviews immediately. There are four main parts in Listing2Speak, which are the website itself, server, AI model and database [13]. The website page of Listing2speak is the platform for users to input their product listing and get reviews. Our system's reviews on aspects of product listing include description, number of description words, number of images, number of videos, number of keywords and sales. After a user enters their product listing page into our web page and chooses their product type, our system will analyze different aspects of their product page and give them feedback on where they should make improvements. Our algorithm adapts to the buying patterns of its customers since the database contains all top 100 products in each amazon product category. Using the datasets collected by web scraping codes, Listing2Speak created a vector regression model in SKlearn machine learning in order to predict and analyze input data. This AI based method provides our system ability to give the most appropriate improvement suggestion for users' product listing in different product categories. Due to the regression model, our system has identified various factors that play into the success of conversions and sales of product listings. Therefore, based on the data relationship of the listing aspects of the product page constructed by the AI model, the server is able to make predictions with input data.

3.3. Website's UI



Figure 4. Screenshot of website

The frontend website is developed using css and HTML [15]. CSS is a program language that defines style structures such as fonts, colors, positions, etc. and is used to describe the way information is formatted and displayed on a web page. CSS styles can be stored directly on HTML pages or in separate style sheet files. This screenshot shows the main UI of Listing2Speak's website. The two red fields below the title are for entering your product listing URL and selecting your product type. After you click on the go button, our backend will start computing and display the content in the bottom half of the image within a few seconds. On the left side of the analysis area below, it shows different aspects that Listing2Speak analyze on of users' input's product listing page. In contrast, in the section on the right, it tells users if they need to improve their product pages and the percentile of their product listing pages in our dataset.

4. EXPERIMENT

In order to verify that our solution can effectively solve problems at different levels and have good user feedback, we decided to select multiple experimental groups and comparison groups for several experiments. For the first experiment, we want to prove that our solution works stable and continuously, so we choose a group size of 100 different products in 4 categories from different retailers and record the revenue changes of them. The goal of the first experiment is to verify if our AI could analyze all products and give the effectively improve suggestions if the AI works good for all categories of the products. Through sampling 10 groups of different targets. Result is collected by statistics if the AI could find and analyze correctly. Experiments have shown that almost all targets in different types tested with the good result which could appearly improve the sell amount of the retailers. Retailer sells Toothpaste has the most improve rates, which means our AI are works more better in analyze pictures. This experiment could explain that the data sets do have a obvious impact on the finding the elements in the sells' website and how to make it better to sell more products. The average increasing sells rate of 4 different categories of products shows below:

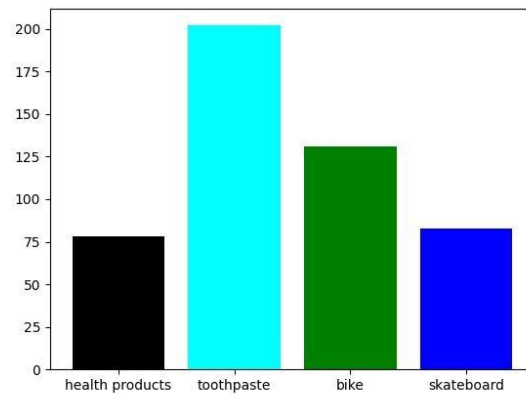


Figure 5. Graph result of experiment

A good user experience is as important as a good product. So a perfect solution should have excellent user experience feedback. In order to prove that our solution has the best user feedback, we specially designed a user experience questionnaire base on the US system usability questionnaire rules. We statistics the feedback result from 100 users, Track the user's data for 10 days using, let them explore freely on the functionality of the web. we divide those users into Four different groups sells health products, toothpaste, bike, skateboard. The goal of the first experiment is to verify if our AI could analyze all products and give the effectively improve suggestions high feedback scores shows high performance We collect the feedback scores form these 4 different group of users and analyze it. Experiments have shown that users who ages who sells toothpaste give the highest result feedback to our web. Which may because of they increasing their revenue a lot by using the web. The experiment graph shows below:

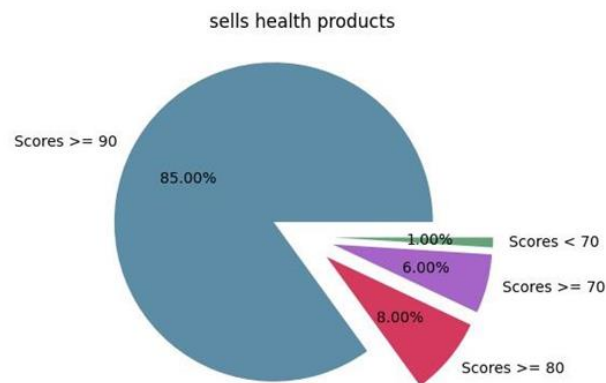


Figure 6. Chart of sells health products

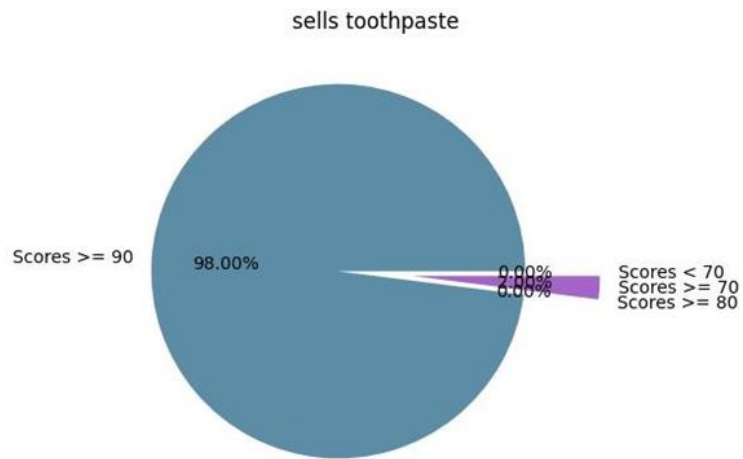


Figure 7. Chart of sells toothpaste

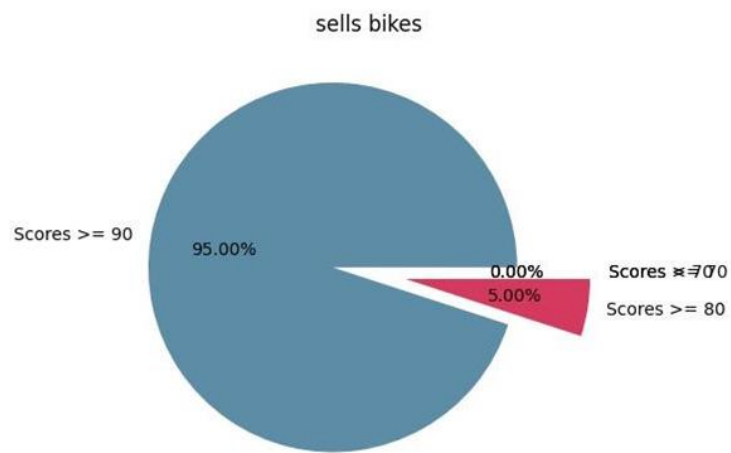


Figure 8. Chart of sells bikes

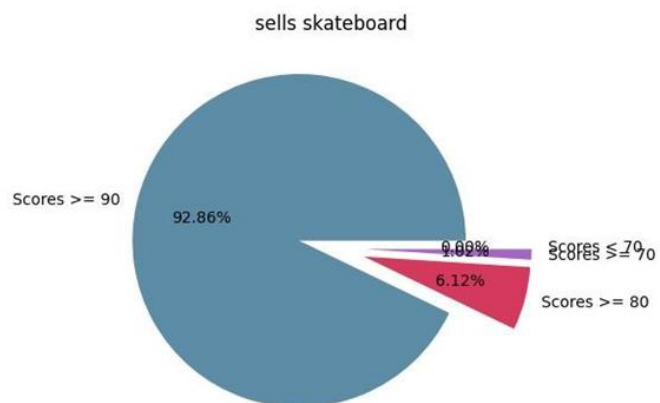


Figure 9. Chart of sells skateboard

5. RELATED WORK

Yan-Kwang C. et al. synthesized the visual attention literature and shelf space allocation theory to build a mathematical planning model with genetic algorithms to find the optimal solution to the focal problem by investigating the effects of display mode (image text vs. plain text) and display format (list vs. array) on users' online shopping performance and attitudes [1]. Their method combined the results of experiments and surveys to provide data support and eventually constructed a mathematical model using functional equations and genetic algorithms. We are using existing data from the Amazon website and linear regression method to create a machine learning model. Yan-Kwang C. et al gives readers a concept how to increase their product's profit, by contrast, Listing2Speak is used to analyze specific user's input and provide ways to improve their sales.

Weiyin Hong et.al's work focuses on two e-commerce page design features: presentation mode and information format [2]. Through their experimental design and observations, they show that the graphic display mode and the list information format are higher than the text-only display mode and the array information format in terms of user performance and attitude, respectively. Their passage is dedicated to proving that combining graphs is better than simple text information. In contrast, Listing2Speak is not concerned with the overall page design of the product page, but rather with the some specific design aspects of the product page that are linked to sales.

Using advanced quantum knowledge and mathematical algorithms, Naoki Nishimura et.al proposed a method to optimize item listings on e-commerce websites, excluding the effect of product popularity and diversity on sales [3]. Their work decomposes the problem by transforming and decomposing it to exclude confounding variables, while Listing2Speak divides the machine learning database into different categories according to Amazon product types to analyze their linear regression separately. The work of Naoki Nishimura et al. has strengths in the mathematics and scientific theory behind their work that support their findings. Listing2Speak uses a machine learning Ai model that focuses more on real-time analytics and provides an improvement solution for individual product pages.

6. CONCLUSIONS

Currently I don't have many project analysis categories, and the database of keyword analysis is not large enough, which leads to most of the analyzed ratings being lower than they would have been. And another limitation about our project is that it can only work when you input an Amazon link which means that you need to distribute your product on Amazon first, only then you can use our e-commerce reviewer. In addition, I want my project to evaluate all the ecommerce products pages, but now I only gathered the database from only the Amazon website. As a result, our project is sometimes not practicable if users don't want to sell their product on Amazon or just test their design demo.

In the future, we are going to gather data from different kinds of e-commerce platforms to make users' choice more diverse. For example, if users want to sell their products through Shopify, they can choose the Shopify category and enter in their Shopify product page's link [14]. We are going to not only expand data for different platforms, but also the keywords dataset which can make our results more accurate.

REFERENCES

- [1] Yan-Kwang Chen, Fei-Rung Chiu, Ciao-Jyun Yang, "An Optimization Model for Product Placement on Product Listing Pages", *Advances in Operations Research*, vol. 2014, Article ID 357693, 9 pages, 2014. <https://doi.org/10.1155/2014/357693>
- [2] Weiyin Hong, James Y.L. Thong, Kar Yan Tam, Designing product listing pages on e-commerce websites: an examination of presentation mode and information format, *International Journal of Human-Computer Studies*, <https://doi.org/10.1016/j.ijhcs.2004.01.006>.
- [3] *Front. Comput. Sci.*, 16 July 2019, Sec. Quantum Engineering and Technology, <https://doi.org/10.3389/fcomp.2019.00002>
- [4] Tian, Yan, and Concetta Stewart. "History of e-commerce." *Encyclopedia of e-commerce, e-government, and mobile commerce*. IGI Global, 2006. 559-564.
- [5] Burt, Steve, and Leigh Sparks. "E-commerce and the retail process: a review." *Journal of Retailing and Consumer Services* 10.5 (2003): 275-286.
- [6] Markovskiy, Ivan, and Florian Dörfler. "Behavioral systems theory in data-driven analysis, signal processing, and control." *Annual Reviews in Control* 52 (2021): 42-64.
- [7] Tran, Lobel Trong Thuy. "Managing the effectiveness of e-commerce platforms in a pandemic." *Journal of Retailing and Consumer Services* 58 (2021): 102287.
- [8] Rajamma, Rajasree K., Audhesh K. Paswan, and Gopala Ganesh. "Services purchased at brick and mortar versus online stores, and shopping motivation." *Journal of Services Marketing* (2007).
- [9] Calixto, Iacer, et al. "Using images to improve machine-translating e-commerce product listings." *Association for Computational Linguistics (ACL)*, 2017.
- [10] Komer, Brent, James Bergstra, and Chris Eliasmith. "Hyperopt-sklearn." *Automated Machine Learning*. Springer, Cham, 2019. 97-111.
- [11] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. "Sentiment analysis on large scale Amazon product reviews." 2018 IEEE international conference on innovative research and development (ICIRD). IEEE, 2018.
- [12] Arena, Fabio, and Giovanni Pau. "An overview of big data analysis." *Bulletin of Electrical Engineering and Informatics* 9.4 (2020): 1646-1653.
- [13] McCarthy, John. "What is artificial intelligence." URL: <http://www-formal.stanford.edu/jmc/whatisai.html> (2004).
- [14] Alfida, P. P., et al. "SHOPIFY – The Smart Shopping Cart." *International Research Journal of Engineering and Technology* 6.6 (2019): 616-620.
- [15] Alawar, Mariam W., and Samy S. Abu-Naser. "CSS-Tutor: An intelligent tutoring system for CSS and HTML." (2017).