Analysis of Ecommerce Platforms Sentiment

Team – 16

Sai Krishna Boinapally

Jatin Raj Thodupunuri

Sunny Sumanth Dodda

Poojitha Achanta

GITHUB-https://github.com/saikrishnaBoinapally/NLP-PROJECT

INTRODUCTION

Consumers in today's world place a great deal of importance on the practice of shopping online since it allows them to reduce the amount of time and effort required to acquire a product. Because of the enormous growth of e-commerce, gathering feedback from customers has become an increasingly important part of identifying their areas of interest and activity. The purpose of doing a sentiment analysis is to ascertain how customers feel about a certain product. This assists other consumers in making judgments about whether or not to purchase the goods. A recommender system that is built on this can offer recommendations to other consumers or show them things that are connected to what they are browsing for. In recent years, sentiment analysis has garnered a considerable deal of interest, as has text categorization based on customer testimonials. Textual reviews, star ratings, and emojis are the many forms that the reviews are presented (Zikang et al., 2020). The shops or service providers may more easily accomplish their goals with the aid of sentiment analysis, which is used to assess the massive amounts of data they collect. Opinion and characteristics based on the information provided on the product's characteristics. A vast quantity of material relating to a certain topic that may be found via social media. People express their thoughts and opinions on social media sites like Twitter, Facebook, and others. After purchasing or utilizing the product, the customer provides feedback about its usefulness. They uploaded a massive amount of information to a variety of different sites. The ability to interpret these reviews provides a huge competitive advantage for businesses, since it enables suppliers to make various judgments concerning the quality of the services or goods being offered. Additionally, the recommender system may be improved with the aid of these evaluations (Yang et al., 2020). We also give information on often purchased items or items that are frequently purchased together, and this is based on the reviews and purchases made by the customer.

Goals and Objectives

The purpose of this study is to provide a system that will be built using a hybrid method that combines context-based engine functionality with stochastic learning. The framework that has been suggested will attempt to create a hybrid recommendation algorithm by combining the many algorithms that are now in use. It will boost performance by overcoming the disadvantages of standard recommendation systems. In addition, the customer sentiment analysis that was carried out for the purpose of this research is an essential instrument for any contemporary company because it enables the business to obtain insights that can be put into action, identify and resolve critical issues with reoccurring patterns that cause customers to feel dissatisfied, strengthen the aspects of a product or service that are responsible for customers' positive emotions, and make decisions that are more data-driven and efficient in general. On a more granular level, the purpose of the customer sentiment analysis carried out on this platform is to provide users with the ability to enhance customer service and, as a result, customer experiences.

Motivation

As the number of online platforms grows quickly, businesses are falling behind and are unable to maintain their competitive advantage over well-funded platforms like Amazon and others. Promoting the application of sentiment analysis is the major factor that propels the platforms that are now the most competitive. Additionally, the majority of the models developed in the earlier study mostly failed to adopt a hybrid technique of stochastic learning, necessitating the use of such a framework in the current study.

Aims and Purposes

The purpose of this study is to provide a system that will be built using a hybrid method that combines context-based engine functionality with stochastic learning. The framework that has been suggested will attempt to create a hybrid recommendation algorithm by combining the many algorithms that are now in use. It will boost performance by overcoming the disadvantages of standard recommendation systems. In addition, the customer sentiment analysis that was carried out for the purpose of this research is an essential instrument for any contemporary company because it enables the business to obtain insights that can be put into action, identify and resolve critical issues with reoccurring patterns that cause customers to feel dissatisfied, strengthen the aspects of a product or service that are responsible for customers' positive emotions, and make decisions that are more data-driven and efficient in general. On a more granular level, the purpose of the customer sentiment analysis carried out on this platform is to provide users with the ability to enhance customer service and, as a result, customer experiences.

Significance

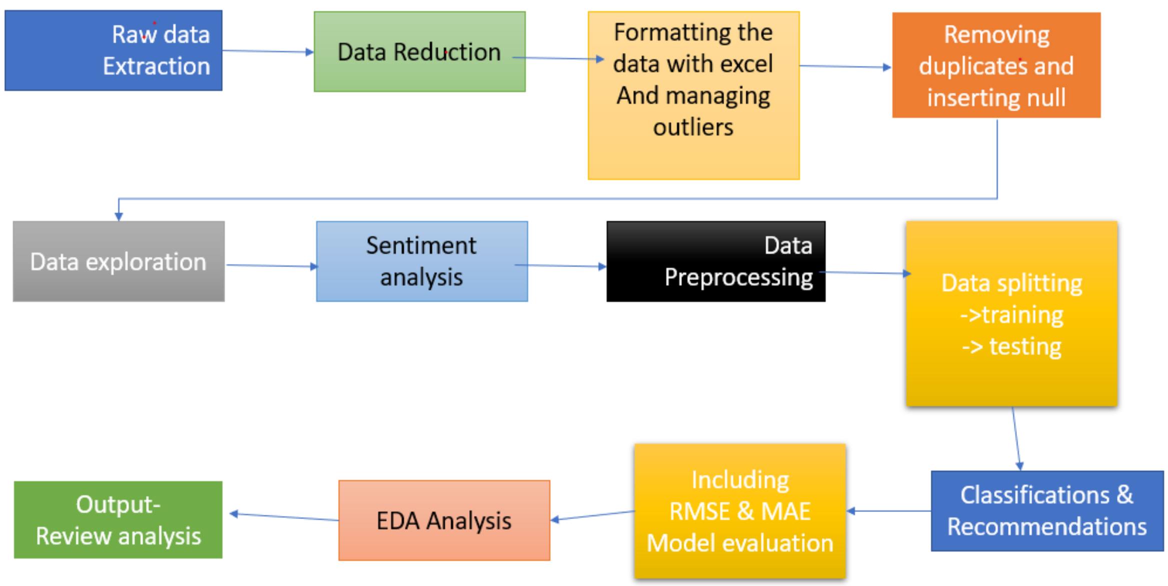
In this day and age of big data, an overwhelming amount of consumer product reviews have been published across various online social media platforms. Consequently, mining the sentiment of customers regarding items can yield significant business knowledge that can enhance the decision-making process of management. Therefore, with the help of the suggested framework for the model, sentiment analysis can be used to investigate a wide range of possibilities, such as the influence sale behavior as well as important brand strategies. Customer analytics, in addition to helping businesses better understand their clients' behaviors, enabling the business to shifts in their clients' requirements. In addition to this, it offers a method for determining which methods of acquiring new consumers and keeping existing ones are successful, as well as which methods are unsuccessful.

Features

The Amazon product reviews that are accessible online are the major source of information for this project. Amazon provides its users with an online option that allows them to review the company's products and services using a star-based scale after making a purchase through the marketplace. Customers also have the option to leave comments, which allow them to more explicitly describe what they took into consideration while assessing the goods. For the sake of this investigation, a data collection consisting of many of these product evaluations will be analyzed using sentiment analysis. The Amazon product reviews area of this website is where the data that will be used in this research will be acquired from. The dataset that was utilized for the research includes features include:

- ✓ reviews.title,
- ✓ brand,
- ✓ reviews.text,
- ✓ categories,
- ✓ primary categories,
- ✓ And the sentiment – contains negative and positive labels.

Block diagram



Background

Businesses that are involved in electronic commerce such as Amazon and Flipkart employ a variety of recommendation algorithms to provide customers various choices. Amazon now employs item-item collaborative filtering, which is capable of scaling to enormous datasets and producing high-quality recommendation systems in real time. This system is a sort of data filtering system that attempts to forecast the "rating" or preferences of the person who is interested in the content being filtered. Amazon's Recommendation System adheres to the notion of creating product-based suggestions. This entails determining the degree to which two items are comparable to one another and then advising individual users on which products are most comparable to those examples (Jin et al., 2013). Researchers' primary interest has always been in developing new metrics for comparing the degree to which two different things are same. However, when it comes to a website such as Amazon, it needs to incorporate more criteria in order to propose things to its visitors, such as the product's quality. Because a product of high quality will almost always have a sizeable number of reviews, we are able to provide suggestions based on both the similarity score and the reviews of individual products (Jiang, 2016). E-commerce websites and other online companies, such as social networking and movie/music rendering sites, now typically include recommendation algorithms as an essential component of their offerings. They have a significant influence on the amount of income that these companies bring in and also provide consumers with benefits by lowering the amount of mental effort required to conduct searches and sort through an excessive amount of data (Hu & Zhang, 2012). A customer's experience may be made more personalized by recommender systems, which analyze a user's interactions with a system and then provide suggestions about other products the user would find helpful.

Dataset

The raw data that we used came from the website of datafinti, which includes a total of 34660 records in its database. There are over 20 columns, some of which include product id, category, sub-category, review title, review text, reviewer name, and other similar information. Raw dataset included a significant number of null values; moreover, both the data and pricing format were incorrect; finally, the dataset had both outliers and noise. The repository is a fairly complete set of customer evaluations and ratings of items sold on Amazon as a result of its creation, which involved scraping millions of Amazon pages. This information may be put to use for a variety of purposes, and it works very well as a foundation for developing a recommendation system. This dataset was compiled using user feedback on Amazon items and contains over 34660 reviews from customers on devices like as the Kindle, Fire TV Stick, and others.

Details design of features

Because the user-book matrix is so sparse, we theorized that we could develop a collaborative filtering model that would work for all 34660 reviews; however, doing so would require a significant amount of time to train, and the resulting embedding would be of poor quality. In this research, we have focused on the two most important columns, reviews and the sentiment.

Analysis

Python was used to trim the raw data since there were more than 20 columns; however, after the reduction, there were only 2 columns remaining for our study. We also use the built-in features of Excel to format a few data and pricing columns so that they have the appropriate format. In Excel, we manage outliers with the help of a built-in function. After that, the null variable was utilized in Python, and duplicates were removed using the filter. The segregated data that we utilized for the analysis are displayed in Figure 1. It has a smaller number of columns compared to the dataset that was first used. In addition, in sentiment analysis it is necessary, prior to engaging in data exploration, to have an understanding of the terms that are most frequently occurring in the evaluations and to eliminate stop words such as "in" and "is." This may be accomplished with the Natural Language Processing (NLP) module for Python known as SpaCy.

```
In [9]: df
Out[9]:
```

	reviews	sentiment
0	This product so far has not disappointed. My c...	5.0
1	great for beginner or experienced person. Boug...	5.0
2	Inexpensive tablet for him to use and learn on...	5.0
3	I've had my Fire HD 8 two weeks now and I love...	4.0
4	I bought this for my grand daughter when she c...	5.0
...	...	...
34655	This is not appreciably faster than any other ...	3.0
34656	Amazon should include this charger with the Ki...	1.0
34657	Love my Kindle Fire but I am really disappoint...	1.0
34658	I was surprised to find it did not come with a...	1.0

Figure 1

Implementation

Despite the numerous other kinds of recommendation systems, such as ones based on quality, classification models, feature recommendation systems, as well as more, the sort of recommendation system that will be the topic of this paper is an overview recommendation system in machine learning, and we will explore how to create it using python code. In the past, recommendations were determined by looking at product patterns, which meant that the product that was being used most frequently was the one that was suggested to practically everyone. Other methods of determining recommendations made use of rating histories.

Jupiter notebook/ Google Colab

For the purpose of this recommender, the Jupiter notebook implementation that Anaconda supplies is utilized. The scalable machine learning library known as Jupiter notebook. The Jupiter notebook is particularly adept at doing computations in an iterative fashion, which enables it to operate quickly.

The Anaconda technique with Weight Regularization is what the Jupiter notebook implements. It is a distributed version of the approach. At the moment, model-based collaborative filtering is supported by Jupiter notebook.

The selection of the model and the hyper-parameters

The main basic parameters that are specified for the recommendation system are the amount of iterations, lambda, as well as rank (the number of latent components). When there was no split in the dataset, the values for rank were set to [2, 5, 10, 20], but they were [8, 10, 20] when there was a split of [0.6, 0.2, 0.2]. The value of lambda might run from [0.001 to 50]. Iterations range from 5 to 20 when there is no split in the dataset, but they are always set to 20 when there is a [0.6, 0.2, 0.2] split in the dataset.

Model Evaluation

A variety of criteria, including RMSE, MAE, Precision, and recall, are utilized in the process of evaluating recommendation engines. For the evaluation of the model, we made use of the RMSE. The RMSE should be as low as possible for the recommendation engine to perform at its best.

Preliminary results

Data preprocessing

EDA analysis

There are 23775 whose products were rated 5 and above with 8541 rated the value of 4. The product with a rate of 2 had the least count.

```
In [69]: df["reviews.rating"].value_counts()

Out[69]: 5.0    23775
         4.0     8541
         3.0     1499
         1.0      410
         2.0      402
         Name: reviews.rating, dtype: int64
```

Figure 2

The negative reviews are rated less than or equal to 3 while the positive ones are ranked above 3.

```
In [11]: count_good=df[df['sentiment'] >3]

         count_bad=df[df['sentiment']<=3]

In [12]: count_bad=count_bad[["reviews","sentiment"]]

In [13]: count_good=count_good[["reviews","sentiment"]]
         count_good.reviews
         count_good.reset_index(inplace=True,drop=True)
         count_bad.reset_index(inplace=True,drop=True)

In [14]: count_bad
```

Out[14]:

	reviews	sentiment
0	Didn't have some of the features I was looking...	2.0
1	i Bought this around black friday for $60 hopi...	1.0
2	I bought this tablet for my 4 year old daughte...	1.0
3	I was hoping to use Google launcher with this ...	3.0
4	The tablet works fine. It is responsive with g...	3.0
...	...	...
2306	This is not appreciably faster than any other ...	3.0
2307	Amazon should include this charger with the Ki...	1.0
2308	Love my Kindle Fire but I am really disappoint...	1.0
2309	I was surprised to find it did not come with a...	1.0
2310	to spite the fact that i have nothing but good...	1.0

2311 rows × 2 columns

Figure 3

Both the negative and the positive reviews are right skewed.

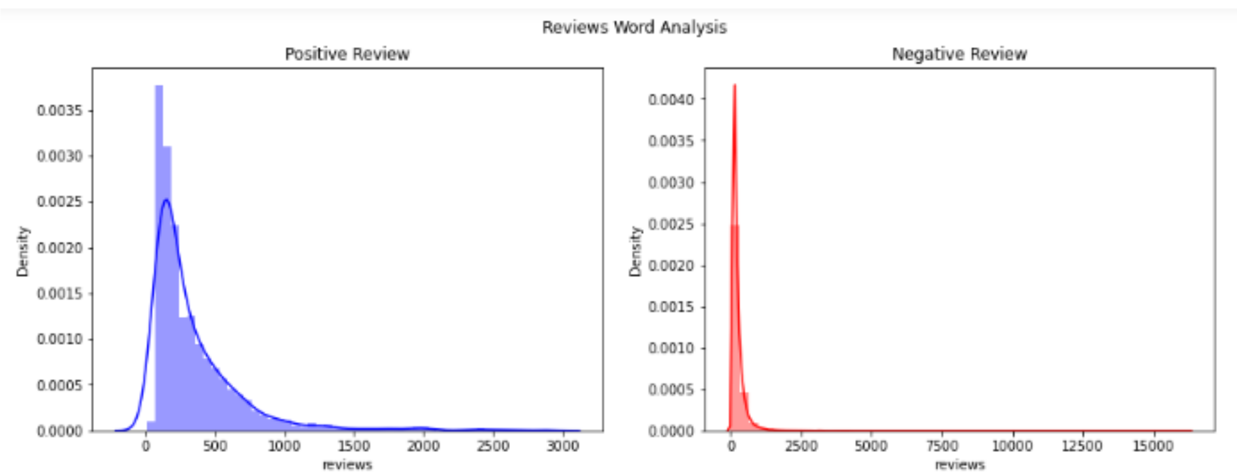


Figure 4

A long-tail normal distribution may be seen in the ratings that were given, regardless of whether they were positive or negative. The total number of positive reviews is 32316, while the number of reviews that are negative is 2311. The lowest number of ratings above three must be at least three, and the maximum number of positive ratings that may be provided is five. According to the data that is provided, a customer provides a positive rating of 4.44 stars on average on Amazon.

Project Management

Work completed.

1. Data collection and comparing with the data which is present in the open websites. **Jatin raj thodupunuri**

2. Data cleaning and formatting the managing outliers and removing analysis. **Sai Krishna Boinapally**

3. Data preprocessing by Sapcy. **Poojitha achanta**

4. Data exploring and classification of models and their selction. **Sunny Sumanth Dodda**

Work to be done.

- Overall executing of code and improving the algorithms. By improving algorithm dataset overfitting and underfitting should be handled. Completing the building of code and execution of code. The desired output is required, and training and testing should be made and the accuracy of algorithm is shown.

References

Hu, J., & Zhang, B. (2012). Product recommendation system. *CS224W Project Report*.
Jiang, J. (2016). Comparative Analysis of E-commerce Recommendation Strategies. In *2016 2nd International Conference on Education Technology, Management and Humanities Science* (pp. 493-496). Atlantis Press.
Jin, Y., Yang, S. B., Rhee, C., & Lee, K. Y. (2013). An Exploratory Study of the Effects of Price Decreases on Online Product Reviews: Focusing on Amazon's Kindle 2.
Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE access*, *8*, 23522-23530.
Zikang, H., Yong, Y., Guofeng, Y., & Xinyu, Z. (2020). Sentiment analysis of agricultural product ecommerce review data based on deep learning. In *2020 International Conference on Internet of Things and Intelligent Applications (ITIA)* (pp. 1-7). IEEE.
P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews", Proceedings of the 40th annual meeting of the association for computational linguistics, pp. 417-424, Dec. 2002, [online] Available: https://arxiv.org/abs/cs/0212032.
P. V. Rajeev and V. S. Rekha, "Recommending products to customers using opinion mining of online product reviews and features", 2015 International Conference on Circuits Power and Computing Technologies (ICCPCT-2015), pp. 1-5, 2015.
"Aamzon review data", Julian McAuley UCSD, 2019, [online] Available: http://jmcauley.ucsd.edu/data/.