Insights Procured From the Recommendation System Based On User Reviews

Shweta Mongia
University of Petroleum and Energy Studies(UPES)
Department of Informatics, School of Computer Science
Energy Acre Building, Bidholi
Dehradun -248007, Uttrakhand, India
Shweta.mongia@yahoo.com

ABSTRACT
With the dawn of the 21st century we saw an unprecedented influx of data. It is only in the past decade that our processing capabilities have caught up to being able to utilize the huge amount of data with several algorithms to be useful. Product reviews and ratings are popular tools to support buying decisions of consumers. These tools are also valuable for online retailers, who use rating systems in order to build trust and reputation in e- commerce. Many online shops offer quantitative ratings, textual reviews or a combination of both. The number of reviews on Amazon has grown significantly over the years. Customers who made purchases on Amazon provide reviews by rating the product from 1 to 5 stars and sharing a text summary of their experience and opinions of the product. This research aims to provide statistical insights into Amazon product reviews, examine their helpfulness in recommending products, and suggest a new way to predict the helpfulness score of the user reviews.
Keywords: *content-based Filtering, collaboratively Filtering, Alternating least squares (ALS)*

1. INTRODUCTION

Personalization is the future of the internet and has achieved tremendous success in industrial applications. Online retailers, such as Netflix and Amazon, for example, offer personalized reviews based on the profile of a customer for additional goods or services. Due to their possible ability to infer the interests of a user from the user background, recent applications such as YouTube, My Google, flip kart, Facebook and Google News have attracted attention [1]. There are various recommendation techniques aimed at recommending products to users for the same purpose[2], and these different techniques can be generalized into three techniques first is Filtering collaboratively, second is Filtering content-based and the last one is Hybrid schemes[3]. Content-based personal recommendation systems [4] are one major personalization issue in the information retrieval community. These systems learn user-specific profiles from user reviews and propose data customized to the interest of each individual user without requiring an explicit question from the user. For these types of systems, learning user profiles is the core issue. Collaborative spark filtering, widely used for recommendation systems, is used [5]. These approaches seek to fill in a user-item association matrix's missing entries. Spark.mllib advocates model-based collaborative filtering, in which a small collection of latent factors that can be used to predict missing entries are identified by users and items. To consider these latent factors, spark.mllib uses the alternating least squares (ALS) algorithm.
Today, some sort of recommendation system (RS) is constantly driven by online consumers, be it for shopping or surfing the internet. RS could be as easy for shopping as the most common item sold on the web, or as advanced as a tailored recommendation based on the history of customer transactions or Collaborative filtering (e.g. the "Consumers who bought this item also bought" from Amazon.com[6].
The objectives of the research are:

1. To deliver Relevant Content to the customers
2. To Increase the number of Items per order
3. To increase the revenue and Customer satisfaction.

The paper is organized into the following sections. Section I conferred with the prefatory phase. Section II explores the related research work. Section III reveals the methodology used. In Section IV, a demonstration of the designated work has been given. Section V concludes with the work and shows the future preview.

2. LITERATURE REVIEW

Sales Recommendation Algorithms are mainly used in Ecommerce Websites like Amazon, Flipkart, etc [7]. In these algorithms, Websites like Amazon have a list of recommended items which are related to customer's interest. With the help of attributes like items viewed, items added to cart, subject interests, demographic data, item purchased and lot more are rated to represent customer's interest [8]. At Amazon, recommendation algorithms are used to personalize and recommend products in online store to each customer [9]. For each customer, store shows products related to customer interests like programming books for Software engineer to latest home appliances to mother. There are mainly three different approaches used for Recommendation such as content based, traditional collaborative filtering or hybrid based etc [10].

In our case, our algorithm will show recommendation based on large data sets related some of attributes related to customer's interest to show quality recommendations [11] Today, some sort of recommendation system (RS) is constantly driven by online consumers, be it for Shopping or surfing the internet.

3. METHODOLOGY

Fig 1 shows the overall methodology used in this study. It consists of data collection, data preparation, data analysis and data visualization phase.

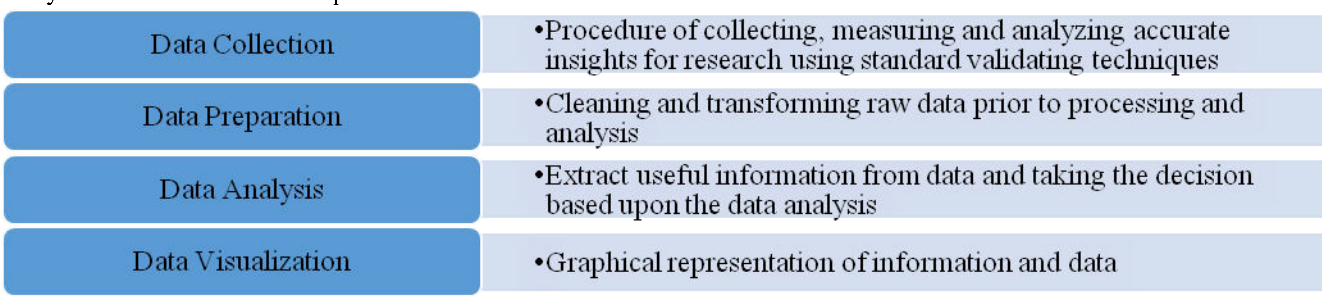

Figure 1: Overall methodology used

3.1 Dataset

Fig 2 shows the dataset used for this study. The dataset attributes like reviewerID (a unique ID given to every reviewer customer), productID (a unique ID given to every product), overall (a overall rating given to every product on the basis of various factors), summary (various keywords which defines the quality of the product) and reviewTime (particular timestamp at which a user gives its review & rating on product) are the attributes which have been used for this research.

reviewerID	productId	overall	summary	reviewTime
AO94DHGC771SJ	528881469	5	Gotta have GPS!	02/06/2013
AMO214LNFCEI4	528881469	1	Very Disappointed	25/11/2010
A3N7T0DY83Y4IG	528881469	3	1st impression	09/09/2010
A1H8PY3QHMQQA0	528881469	2	Great grafics, POOR GPS	24/11/2010
A24EV6RXELQZ63	528881469	1	Major issues, only excuses for support	29/09/2011
A2JXAZZI9PHK9Z	594451647	5	HDMI Nook adapter cable	03/01/2014
A2P5U7BDKKT7FW	594451647	2	Cheap proprietary scam	27/04/2014

Figure 1. Dataset attributes

3.2 Process Flow

1. Dataset has been collected
2. Clean the dataset means removing the null values and redundant data.
3. Convert the categorical data into numerical.
4. Apply the data scaling.
5. Divide the data into train and test.
6. Apply the Alternating Least Squares (ALS) Spark ML[12]. Apache Spark ML implements alternating least squares (ALS) for collaborative filtering, a very popular algorithm for making recommendations. ALS

recommender is a matrix factorization algorithm that uses Alternating Least Squares with Weighted Lamda-Regularization (ALS-WR).

7. It factors the user to item matrix A into the user-to-feature matrix U and the item-to-feature matrix M and calculate the Root means Square Error.

3.3 Tools and Techniques

Alternating Least Squares (ALS) Model works on Matrix Factorization. In the case of collaborative filtering, matrix factorization algorithms work by decomposing the reviewer-product interaction matrix into the product of two lower dimensionality rectangular matrices. One matrix can be observed as the user matrix where rows represent users, and columns are latent factors. The other matrix is the item matrix where rows are latent factors, and columns represent items. It uses L2 Regularization. Its training routine is bit different, as it first of all hold reviewed Matrix and runs the Gradient Descent on Product Matrix, then it holds the Product Matrix and runs Gradient Descent on the Reviewer Matrix.

4. RESULTS AND DISCUSSION

After importing the data the customers are being segregated based on the ratings which they have given as shown in fig 3.

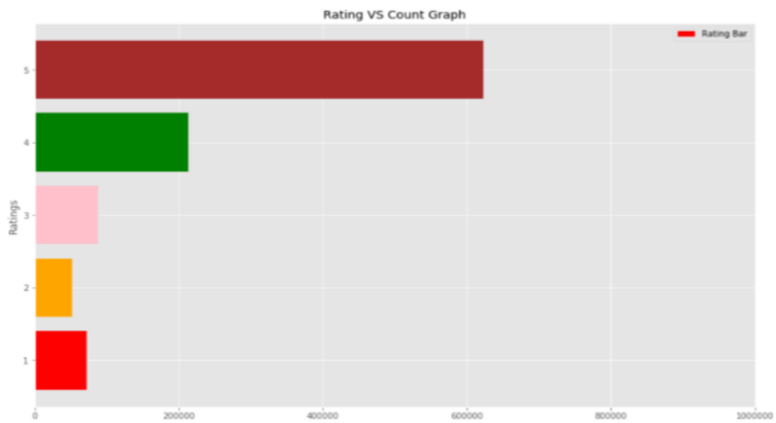

Figure 3: Segregation of customers based on ratings

As indicated by fig 4, pie chart shows rating percentage of a particular product given by users. As shown above, the pie chart of this product tells us that 59% users have given 5 star rating, 20% users have given 4 star rating, and So on. This pie chart helps to overall evaluate the quality of product based on rating given by users.

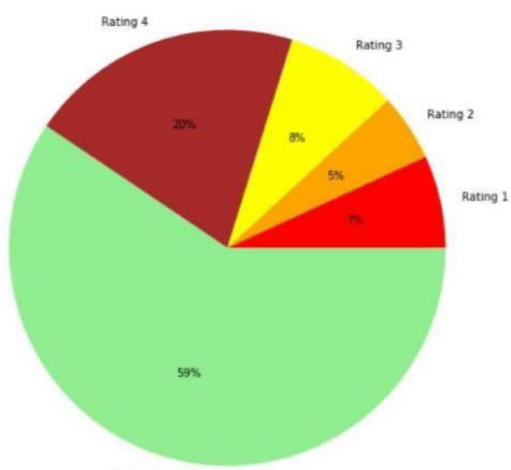

Figure 4: Pie chart of the rating count

This graph represented by fig 5 shows the Range of Rating vs Sales Plot. If Rating >1 & <=2 it shows in green color, if Rating >2 & <=3 it shows in Maroon color, if Rating >3 &<=4 it shows in yellow colour and So on. This graph also shows if a product have higher Average rating then sales frequency of that product is also higher as compared to product which have less average rating.

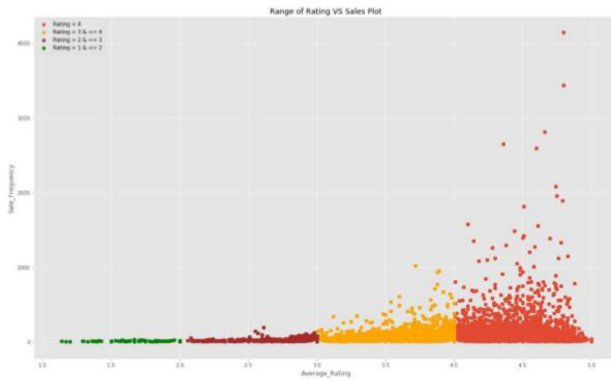

Figure 5: Range of rating Vs Sales Plot

Graph represented by fig 6 shows the average rating which is given to a particular product. On the y axis it shows the sales frequency of that particular product and on the x axis the average rating.

Figure 6: Average rating Vs Sales Plot

Graph in fig 7 plots unique reviewerId vs Total ratings. This graph shows count of total rating given by every unique Reviewer. Suppose a unique reviewer ‘AZ139NBM9’ have given of total 10 rating on different products on website while another reviewer ‘AZ139NKLP’ have given total 20 rating on different products on website. So This plot shows the the reviewer ‘AZ139NKLP’ had engaged more on website as compared to user‘AZ139NBM9’.

Figure 7: unique reviewerId vs Total ratings

This table 1 has a additional column of Average_Rating given to a particular Product_Id.

Table1: Additional column Average_Rating

ProductID	ReviewerID	ReviewTime	Average_Rating	ProductID_Index	ReviewerID_Index
B003ES5ZUU	A28Z618U150FUN	09-07-2014	4.8	0.0	42819.0
B003ES5ZUU	A38R82NFXD83Y3	29-04-2013	4.8	0.0	56945.0
B003ES5ZUU	A2ALUVGJ3WMWW4	28-09-2013	4.8	0.0	43461.0
B003ES5ZUU	A1AZOZ63I440RR	23-06-2014	4.8	0.0	29243.0
B003ES5ZUU	A1BEWP50THJDNP	28-01-2014	4.8	0.0	29410.0
B003ES5ZUU	A19PQP53WON0LR	07-07-2013	4.8	0.0	28731.0
B003ES5ZUU	AUXUZB6JGFLMM	24-02-2014	4.8	0.0	77065.0
B003ES5ZUU	A2W0GY64CJSV5D	03-01-2013	4.8	0.0	1892.0
B003ES5ZUU	A35NRB38PPMDWM	06-01-2014	4.8	0.0	5538.0
B003ES5ZUU	A3ICJ5EDDE3G98	10-11-2013	4.8	0.0	19088.0
B003ES5ZUU	A206WRB4UEBNOS	22-01-2013	4.8	0.0	12265.0
B003ES5ZUU	AACH5Q7O04C2K	08-03-2014	4.8	0.0	21636.0
B003ES5ZUU	A25GVTHHZL2XBP	26-01-2014	4.8	0.0	41443.0
B003ES5ZUU	AS4P74KQY9WR4	10-05-2013	4.8	0.0	1080.0
B003ES5ZUU	A2CVQF1FJ5I33Q	01-01-2013	4.8	0.0	1665.0
B003ES5ZUU	A2AFFIW9KTDB8O	11-01-2014	4.8	0.0	43387.0
B003ES5ZUU	A1UKWA1P58XG08	17-01-2012	4.8	0.0	11511.0

The table 2 given below above is final output. As shown below, the attribute ProductID_Index is the unique product ID which will be recommended to customer given in 'Recommended_To_Reviewers' list.

Table2: Recommendation table

Average_Rating	ProductID_Index	ReviewerID_Index	prediction
3.72	27.0	156.0	3.2792048
3.72	27.0	324.0	3.3408682
3.72	27.0	397.0	3.4857466
3.72	27.0	649.0	3.3400931
3.72	27.0	683.0	3.306909
3.72	27.0	1064.0	3.3817198
3.72	27.0	1120.0	3.3398416
3.72	27.0	1565.0	3.3131092
3.72	27.0	2032.0	3.2459548
3.72	27.0	2045.0	3.347926
3.72	27.0	2126.0	3.3368309
3.72	27.0	2210.0	3.2125227
3.72	27.0	3685.0	3.36837
3.72	27.0	4056.0	3.4694424
3.72	27.0	4466.0	3.3478358
3.72	27.0	5485.0	3.224565
3.72	27.0	5717.0	3.2979357

5. CONCLUSION

After applying the model we found the importance of relevant content which helps ecommerce websites like ebay, amazon, flipkart, etc drives traffic to their website. Recommendation of products based on customer's interest and previous search queries also help in Increase number of Items per order. We concluded that predictive analysis of each product based on user rating is important from the perspective of customer satisfaction and increment in revenue.

REFERENCES

1. Singh, Pradeep & Dutta Pramanik, Pijush & Dey, Avick & Choudhury, Prasenjit. (2021). Recommender Systems: An Overview, Research Trends, and Future Directions. International Journal of Business and Systems Research. 15. 14–52.
2. Joeran Beel, Bela Gipp, Stefan Langer, Corinna Breitinger (2016), Research-paper recommender systems: a literature survey, International Journal on Digital Libraries ; 17 (2016), 4. - S. 305-338 https://dx.doi.org/10.1007/s00799-015-0156-0
3. Francesco Ricci and Lior Rokach and Bracha Shapira, Introduction to Recommender Systems Handbook, Recommender Systems Handbook, Springer, 2011.
4. Pazzani M.J., Billsus D. (2007) Content-Based Recommendation Systems. In: Brusilovsky P., Kobsa A., Nejdl W. (eds) The Adaptive Web. Lecture Notes in Computer Science, vol 4321. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_10
5. Schafer, Ben & J, Ben & Frankowski, Dan & Dan, & Herlocker, & Jon, & Shilad, & Sen, Shilad. (2007). Collaborative Filtering Recommender Systems.
6. G. Linden, B. Smith and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," in IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, Jan.-Feb. 2003, doi: 10.1109/MIC.2003.1167344.

7. Pegah Malekpour Alamdari, Nima Jafari Navimipour, Mehdi Hosseinzadeh, Ali Asghar Safaei, And Aso Darwesh, (2020), A Systematic Study on the Recommender Systems in the E-Commerce, IEEE Access, Vol 8, 2020.
8. Xiao, Bo, and Izak Benbasat. "E-commerce product recommendation agents: use, characteristics, and impact." MIS quarterly 31.1 (2007): 137- 209.
9. Smith, Brent & Linden, Greg. (2017). Two Decades of Recommender Systems at Amazon.com. IEEE Internet Computing. 21. 12-18. 10.1109/MIC.2017.72.
10. M. V. Murali, T. G. Vishnu and N. Victor, "A Collaborative Filtering based Recommender System for Suggesting New Trends in Any Domain of Research," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019, pp. 550-553, doi: 10.1109/ICACCS.2019.8728409.
11. Chen, Pei-Yu, Shin-yi Wu, and Jungsun Yoon. "The impact of online recommendations and consumer feedback on sales." ICIS 2004 Proceedings (2004): 58.
12. Jung-Bin Li; Szu-Yin Lin; Yu-Hsiang Hsu; Ying-Chu Huang,(2020), An empirical study of alternating least squares collaborative filtering recommendation for Movie lens on Apache Hadoop and Spark, International Journal of Grid and Utility Computing, 2020 Vol.11 No.5, pp.674 – 682.