Recommendation System Using Product Rank Algorithm For E-Commerce

Urvesh Rathod[1], Prof. Aruna Pavate[2], Vaibhav Patil[3]
[1]*(Computer Engineering, Atharva College of Engineering/Mumbai University, India)*
[2]*(Computer Engineering, Atharva College of Engineering/Mumbai University, India)*
[3]*(Computer Engineering, Atharva College of Engineering/Mumbai University, India)*

Abstract: *Evolution of E-commerce in today's scenario is deliberately growing, and in order to boost client's sales and stay ahead of the competition, it is important to get on with online shopping trends and understand customer's needs. Shopping websites have shown atremendous rise in product availability on the website in past few years. This arises a problem of exploding databases. Websites use many data mining algorithms and recommendation system but still, users might or might not get the best product for their search which results in declination of customer experience. Minimizing the search results to optimality and integrating products from multiple websites can economize time and customer's experience. Moreover, no shopping website has price tracking and prediction algorithm for thecustomer. Our System tends to provide customers with rich and expensive products from all available shopping websites at thecheapest price of their choice without using manual filters, with the help of web scraping and Product Rank algorithm. With the increasing profits due to sales, we must also manifest the loss incurred due to bargain with the customers. Our System will also be able to track down dead pages and price errors that can prove as a good forensic system for different shopping websites.*

Keywords : *E-commerce,Web spider, Crawler, Product Rank Algorithm (PDR), Recommendation System*

I. Introduction

Almost every human being is well known of the word E-commerce, as it becomes anindivisible part of our life. E-commerce is most advancing and progressive platform where there are millions of products available to billions of customers. Every second thousand of products is purchased on e-commerce sites. In India e-commerce is growing faster than other countries as the annual rate of purchasing is 51% in India and it is increasing day by day. [1][2]

We can see that every festival season there are many sales, offers,and discounts on products of e-commerce sites, which leads to customer attraction and customer growth. Also discounts given on the product to increase demand for that particular product. In the season of festival, sale, offers customers are not able to find their desired product from thousands of varieties of products at best quality and cheap price. Sometimes it is aserver issue or maybe traffic issues at the site. A common person cannot find the product from so many different products. Filters may work at that time. [4]

Our System will work on different types of e-commerce websites. As acrawler, it will extract all products from e-commerce sites and generate alist of products. It filters all products on the basis of best quality and price. Great discounted products will be listed first, so it becomes easy to find out adesired product with thebest quality at best price. At a time,Our System will give alist of featured products from multiple e-commerce websites like Flipkart, Amazon, Snapdeal, and Paytm. [4]

II. Implementation

Our System consists of web scraping module and data mining module. It was implemented using Python on PyCharm IDE. The user enters his/her desired product into the search bar and waits for output. Along with the Name of the product, theuser is also asked to enter how much discount percentage is desired and what is his/her maximum budget. Also, these data are stored in the activity tracker database in case if the same user visits again and searches for anew product, Our System will act as a recommendation system too. After that, all the E-Commerce websites associated with that product will be scraped and stored in the database. Our System will then use Product Rank Algorithm to rank products accordingly and also will try to find thesimilar product using Recommendation System. The whole process nearly requires few seconds if the user is having at least speed of 1Mbps internet connection and a system with 1 GHz Processor.

A. Dataset:

Using Web-Scraping technique, data of around 8000 entries had been collected and stored in MongoDB database. The dataset contains following attributes:

Table 1: Parameters considered for our system.

Name of the Product	Category of the Product	Deal Price	Normal Price	Discount	Popularity	Customer Review

B. Steps of the system:

Step 1: User enters the search input and minimum discount on aproduct with its budget
Step 2: The entered input will be saved in activity tracker database file and will be used as a parameter in web scraping.
Step 3: As per user's input, E-Commerce websites are scraped and stored in the database.
Step 4: Applying Product Rank Algorithm and Recommendation Algorithm to therank product according to user's search parameter.
Step 5: Displaying the filtered products to theuser

C. Product Rank Algorithm:

After Scraping products from the websites, it is not necessary that products are filtered according to user's search parameter so our system performs 2-layer filtering for theuser.

The Initial value for every parameter will be given a value of 1/n (n = no. of parameters) and based on it we rank our scraped products. In the end, the product with thehighest rank will be displayed first to the user.[4][5]

We assume there are Parameter $A_1 \dots A_n$ who has products $T_1 \dots T_n$ which point to it. The parameter d is the damping factor which can be set between 0 and 1. Also, $C(A_1) \dots C(A_n)$ is defined as the number of links outgoing from Parameters. Then the Product Rank (PDR) of Product $T^{(i)}$ (where i is the desired product) is given as follows:

$$PDR(T^{(i)}) = (1-d) + d(PDR(A_1)/C(A_1) + \dots . + PDR(A_n)/C(A_n)) \dots \dots \dots \dots \dots \dots \dots \dots \mathbf{(1)}$$

Where,
$PDR(A_{(i)})$= Product Rank of Parameter A at ith Position
$PDR(T^{(i)})$ = Product Rank of Product T at ith Position associated with Parameters
$PDRC(A_{(i)})$ = Total No. of outgoing links from Parameter Aat ith Position
d = Damping Factor. Here we assume d = 0.85

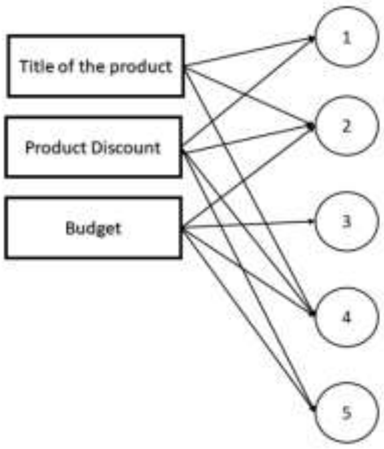


Fig.1 consists of parameters in the boxes and nodes represent the product associated with that parameters

Consider the following example shown in Fig. 1 which consistsof three different parameters listed in the boxes and all the circle represents products.

The total no. of incoming links to a specific product decides the Product Rank of the following product. If there does not exist a link from a parameter to a product it will be considered as zero.

Initial Assumptions: PDR (Title of the product) = 0.33333
PDR (Product Discount) = 0.33333

PDR (Budget) = 0.33333

Table 2: Calculation of Product Rank for Products shown in Fig. 1

Products	X = PDR(A_1)/C(A_1) + …. + PDR(A_n)/C(A_n)	PDR($T^{(i)}$) = (1-d) + d(X)	Product Rank
Product 1	0.11	0.2435	2
Product 2	0.2776	0.38596	5
Product 3	0.083	0.22055	1
Product 4	0.2776	0.38596	5
Product 5	0.1667	0.2917	3

Here Product 2 and Product 4 have highest Product Rank and therefore Product 2 and Product 4 will be displayed first. We can also adjust the Product Rank values of parameters in case one parameter holds more weight than the other parameter.

D. Recommendation Algorithm:

After scraping all the desired products, the parameters from activity tracker database will be used to recommend more products to the user.

Consider an example, where there are 4 users named Alice, Bob, Chris and Daniel have thecommon interest in playing games. There are multiple games which are based on Strategy or Action. The table demonstrates the liking of the game from 0 to 5 rating of an individual person.

Table 3: Example dataset for recommendation system

Games	Category	Alice	Bob	Chris	Daniel
Clash of Clans	Strategy	5	5	0	0
Dota 2	Strategy	4	?	1	1
Dark Souls	Adventure	?	1	5	4
Darksiders	Adventure	1	1	5	5

In order to predict the values which are not available, we use content-based recommendation System to identify the unknown values.

Here n_u = no. of users, n_g = no. of games, r(i,j) = 1 if user has rated gate i, $y^{(i,j)}$ = rating given by user j to game i (defined only if r(i,j) = 1). Let x1 and x2 be the degree to measure a game and x0 = 1.

Table 4: Adding Slack Variable X1 and X2 as the degree to measure games

Games	Alice	Bob	Chris	Daniel	X1(Strategy)	X2(Action)
Clash of Clan	5	5	0	0	1	0.01
Dota 2	4	?	1	1	0.9	0
Darks Souls	?	1	5	4	0.1	1
Darksiders	1	1	5	5	0	0.9

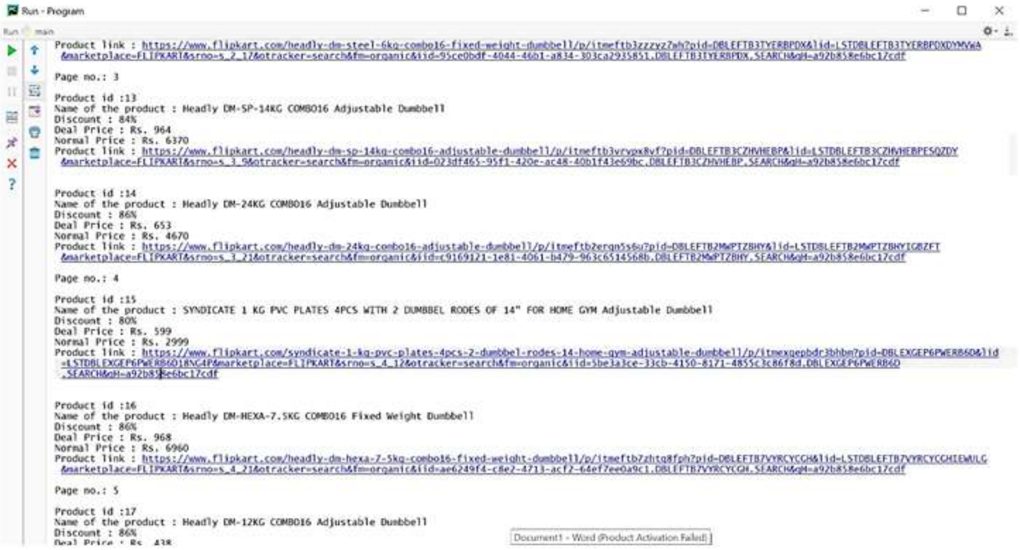
Prediction (i, j) = $(Ø^{(j)})^{T} * X^{(i)}$……………………….**(2)**

Where,
$X^{(i)}$ = Vector consisting of x0, x1, x2 elements.
$Ø^{(j)}$ = Prediction learner variable for user j and for each user j, $Ø^{(j)} = R^3$.
$(Ø^{(j)})^{T}$ = Transpose of $Ø^{(j)}$.
Prediction (i, j) = Predict user j as rating game i with $(Ø^{(j)})^{T} * X^{(i)}$ stars.

For the above example, $X^{(2)} = [1\ 0.9\ 0]_{1x3}$ and $Ø^{(2)} = [0\ 5\ 0]_{3x1}$.
Therefore, Prediction (2, 2) = $(Ø^{(j)})^{T} * X^{(i)}$ = 0.9 * 5 = 4.5

Here we have used simple content based recommendation system but recently there are many powerful recommendation systems available which can make Our System more robust. The main motive of implementing Content Based Recommendation system is to improve Our Systems accuracy once user entered data is stored in our database. It is also helpful to identify customer's interest and amplify business models. [9] [10]

III. Result And Analysis



```
Product link : https://www.flipkart.com/headly-dm-steel-6kg-combo16-fixed-weight-dumbbell/p/itmeftb3zzzyz7wh?pid=DBLEFTB3TYERBPDX&lid=LSTDBLEFTB3TYERBPDXDYMVWA
 &marketplace=FLIPKART&srno=s_2_17&otracker=search&fm=organic&iid=95ce0bdf-4044-46b1-a834-303ca2935851.DBLEFTB3TYERBPDX.SEARCH&qH=a92b858e6bc17cdf

Page no.: 3

Product id :13
Name of the product : Headly DM-SP-14KG COMBO16 Adjustable Dumbbell
Discount : 84%
Deal Price : Rs. 964
Normal Price : Rs. 6370
Product link : https://www.flipkart.com/headly-dm-sp-14kg-combo16-adjustable-dumbbell/p/itmeftb3vrvpx8vf?pid=DBLEFTB3CZHVHEBP&lid=LSTDBLEFTB3CZHVHEBPESQZDY
 &marketplace=FLIPKART&srno=s_3_9&otracker=search&fm=organic&iid=023df465-95f1-420e-ac48-40b1f43e69bc.DBLEFTB3CZHVHEBP.SEARCH&qH=a92b858e6bc17cdf

Product id :14
Name of the product : Headly DM-24KG COMBO16 Adjustable Dumbbell
Discount : 86%
Deal Price : Rs. 653
Normal Price : Rs. 4670
Product link : https://www.flipkart.com/headly-dm-24kg-combo16-adjustable-dumbbell/p/itmeftb2ergn5s6u?pid=DBLEFTB2MWPTZBHY&lid=LSTDBLEFTB2MWPTZBHYIGBZFT
 &marketplace=FLIPKART&srno=s_3_21&otracker=search&fm=organic&iid=c9169121-1e81-4061-b479-963c6514568b.DBLEFTB2MWPTZBHY.SEARCH&qH=a92b858e6bc17cdf

Page no.: 4

Product id :15
Name of the product : SYNDICATE 1 KG PVC PLATES 4PCS WITH 2 DUMBBEL RODES OF 14" FOR HOME GYM Adjustable Dumbbell
Discount : 80%
Deal Price : Rs. 599
Normal Price : Rs. 2999
Product link : https://www.flipkart.com/syndicate-1-kg-pvc-plates-4pcs-2-dumbbel-rodes-14-home-gym-adjustable-dumbbell/p/itmexgepbdr3bhbm?pid=DBLEXGEP6PWERB6D&lid
 =LSTDBLEXGEP6PWERB6D18NG4P&marketplace=FLIPKART&srno=s_4_12&otracker=search&fm=organic&iid=5be3a3ce-33cb-4150-8171-4855c3c86f8d.DBLEXGEP6PWERB6D
 .SEARCH&qH=a92b858e6bc17cdf


Product id :16
Name of the product : Headly DM-HEXA-7.5KG COMBO16 Fixed Weight Dumbbell
Discount : 86%
Deal Price : Rs. 968
Normal Price : Rs. 6960
Product link : https://www.flipkart.com/headly-dm-hexa-7-5kg-combo16-fixed-weight-dumbbell/p/itmeftb7zhtq8fph?pid=DBLEFTB7VYRCYCGH&lid=LSTDBLEFTB7VYRCYCGHIEWULG
 &marketplace=FLIPKART&srno=s_4_21&otracker=search&fm=organic&iid=ae6249f4-c8e2-4713-acf2-64ef7ee0a9c1.DBLEFTB7VYRCYCGH.SEARCH&qH=a92b858e6bc17cdf

Page no.: 5

Product id :17
Name of the product : Headly DM-12KG COMBO16 Adjustable Dumbbell
Discount : 86%
Deal Price : Rs. 438
```

Fig. 2 shows the output of Our System in which user has passed parameter Name of the product as dumbbells, discount as 80% and budget as 1000. Our System filters out all the details of the product and lists them as shown. The Fig. 2 consists of Page No. which illustrates on which page no. the item is listed on Flipkart, Name of the Product, how much is the discount, Deal Price of the Product, Normal Price of the product and Link for the product.

Using above discussed algorithm, the expected result is shown in Fig. 2.The user has given input parameter Name of the product as dumbbells, discount as 80% and budget as 1000. So Our System lists all the products with these parameters as true and ranks accordingly. For an instance at product id = 15, the discount is 80% and budget is less than INR 1000 which both matches the user’s expectation.

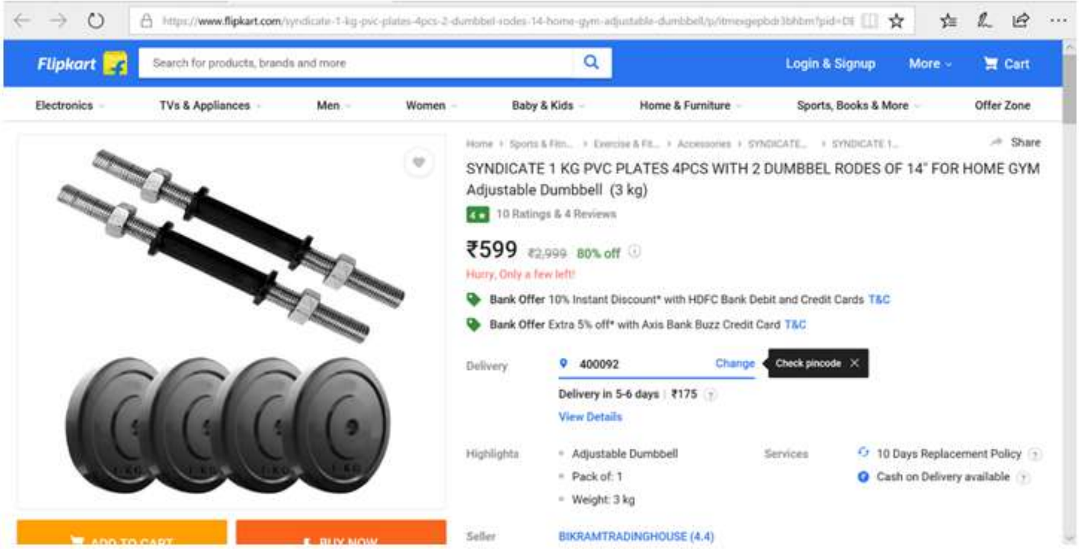
A. Accuracy and Test Results:

The purpose of this proposed system is to provide accurate results based on Product Rank algorithm. The overall accuracy of PageRank algorithm (PR) with KNN classifier is in between 78% - 81% and Theme Weight and Bayesian Page Rank Algorithm (TW BPR)is in between 80% - 85% tested for over 15 different themes and 500 different Web pages. [8]

Out of the 8000 entries of dataset collected, 3000 entries were used for training the algorithm. Rest 5000 entries were used to test the algorithm. 4000 entries from 5000 remaining entries were found accurate while rest 1000 entries gave false results/error pages.

Therefore, Accuracy of our System $= \frac{\text{No.of positive results} + \text{No.of negative results}}{\text{Total entries used}} = \frac{5000+1000}{8000} * 100 = 75\%$

Our System was tested on over 6 different E-Commerce websites on over 8000 different products for testing accuracy results.


https://www.flipkart.com/syndicate-1-kg-pvc-plates-4pcs-2-dumbbel-rodes-14-home-gym-adjustable-dumbbell/p/itmexgepbdr3bhbm?pid=DI
Flipkart
Search for products, brands and more
Login & Signup
More
Cart
Electronics
TVs & Appliances
Men
Women
Baby & Kids
Home & Furniture
Sports, Books & More
Offer Zone
Home › Sports & Fitn... › Exercise & Fit... › Accessories › SYNDICATE... › SYNDICATE 1...
Share
ADD TO CART
BUY NOW

SYNDICATE 1 KG PVC PLATES 4PCS WITH 2 DUMBBEL RODES OF 14" FOR HOME GYM Adjustable Dumbbell (3 kg)

4★ 10 Ratings & 4 Reviews

₹599 ~~₹2,999~~ 80% off

Hurry, Only a few left!

Bank Offer 10% Instant Discount* with HDFC Bank Debit and Credit Cards T&C

Bank Offer Extra 5% off* with Axis Bank Buzz Credit Card T&C

Delivery: 400092 Change | Check pincode ✕

Delivery in 5-6 days | ₹175

View Details

Highlights:
- Adjustable Dumbbell
- Pack of: 1
- Weight: 3 kg

Services:
- 10 Days Replacement Policy
- Cash on Delivery available

Seller: BIKRAMTRADINGHOUSE (4.4)

Fig. 3 is used to prove the accuracy of one of the search results shown in Fig. 2where the user's input is dumbbells with minimum 80% discount and price of the product must be less than INR 1000. The E-Commerce website used here is https://www.flipkart.com.

B. Observations:

In this experiment we have used Fig. 2 and Fig. 3 to demonstrate our test results to determine the accuracy but we also need to observe that is the price of the product righteous? In this case, we can notice that price of the product, despite giving 80% discount, is a bit high as compared to offline stores and yet consumers buy it assuming it is a steal deal. Moreover, it also happens sometimes that due to wrongly entered price by third party seller, E-Commerce websites recompensate consumers justifying price errors/glitches. [11] All the above cases can easily be detected and precautionary measures can be undertaken which results in minimizing intangible losses.

IV. Conclusion

In this Paper, we showed that Our System is very optimal, powerful and accurate which can improve customers experience on E-Commerce websites by boosting up the sales. Moreover, Our System achieved success in finding out expensive and quality products at cheapest rates which was the main motive of our experiment. Consumers no longer need to wait for festive seasons or any occasion to purchase. It has also been observed that during any sale, customers are surfing on E-Commerce website for any steal deal which increases server load and results in crashing the webpage. Our System will divert all that traffic for few seconds on our server i.e decentralizing the traffic to improve efficiency of E-Commerce websites. Once Our System is live as a centralized E-Commerce website, limitations might only get restricted to web scraping. Since E-Commerce are changing their HTML code periodically, some minor changes are must in Our System. Not only Our System is beneficial to consumers but it is also help E-Commerce websites to track down dead pages and price errors immediately and in order to sustain and constantly develop in a competitive era it is necessary to undertake as much precaution measures as one can. Considering all the above factors, it can also be concluded that existence of Our System does not exit and it is might be a wise decision to implementOur System in near future.

A. Future Scope:

In near time, this system will also be integrated with price prediction graphs which will help customer identify the nature of the price of the product during festive events.

References

Websites:

[1] https://www.export.gov/article?id=India-e-Commerce Accessed on (08th August 2017)
[2] https://www.ibef.org/industry/ecommerce.aspx Accessed on (07th December 2017)
[3] Buyhatke- https://compare.buyhatke.com/ 29 January 2017

Journal Papers:

[4] Urvesh Rathod, Aruna Pavate, Vaibhav Patil, "Smart Search Engine using Product Rank Algorithm for E-Commerce". *2018 3rd International Conference for Convergence in Technology (I2CT)*, pp.1-5, 2018.

[5] Ashish Jain, Rajeev Sharma, Gireesh Dixit, Varsha Tomar," Page Ranking Algorithms in Web Mining, Limitations of Existing methods and a New Method for Indexing Web Pages", *International Conference on Communication Systems and Network Technologies, 978-0-7695-4958-3/13 IEEE DOI 10.1109/CSNT.2013.137*, pp.640-645, 2013.

[6] Dr. Daya Gupta, Devika Singh," User Preference Based Page Ranking Algorithm". *International Conference on Computing, Communication, and Automation (ICCCA2016)*, Department of Computer Science and Engineering, Delhi, ICCCA, pp.166-171, 2016.

[7] Lissa Rodrigues, Shree Jaswal, "Hybrid Model for Improvised Page Ranking". 2015, *International Conference on Control, Instrumentation, Communication and Computational Technologies (lCCICCT)*, pp.466-469. 2015.

[8] Yan Tang(&), Lei Wei, Wangsong Wang, and Pengcheng Xuan, "Towards Intelligent Web Crawling – A Theme Weight and Bayesian Page Rank Based Approach", *Springer International Publishing AG 2017 A. Bouguettaya et al. (Eds.): WISE 2017, Part II, LNCS 10570, DOI: 10.1007/978-3-319-68786-5_16*, pp. 192–201, 2017.

[9] Pooja Vashisth,Punam Bedi "Internet based personalized recommender system" *2011 World Congress on Information and Communication Technologies 978-1- 4673-0126- 8/11/$26.00 c 2011 IEEE*, pp.245-250. 2011.

[10] Jieun Son, Seoung Bum Kim, "Content-based filtering for recommendation systems using multiattribute networks", *0957-4174/© 2017 Elsevier Ltd*, pp.404-412, 2017

[11] JunBo Xia, "E-commerce Product Recommendation Method based on Collaborative Filtering technology",*2016 International Conference on Smart Grid and Electrical Automation*, pp.90-93 2016.

[12] Benjamin Groebner, Oops! The Legal Consequences of and Solutions to Online Pricing Errors, 1 Shidler J. L. Com. & Tech. 2 (May 26, 2004), pp.1-6, 2004.

[13] Dai Wei,PengGeng,Liu Ying (Univ. of Chinese Acad. of Sci., Beijing, China) " A prediction study on e-commerce sales based on structure time series model and web search data". *2014 26th Chinese Control and Decision Conference (CCDC*), pp.5346- 5351, 2014.