IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

# Product Recommendation Based on Customer Behaviour using Data Mining

**Sourabh Joshi[1], Pranav Phate[2], Naman Jain[3], Abhijeet R. Raipurkar[4]**

Computer Science Department, RCOEM, Nagpur[1,2,3,4]

**Abstract:** Product Recommendation involves suggesting prospective customers to purchase additional products by analyzing their purchase history, similarity of buying pattern with respect to other customers, and items in their shopping basket. This process of recommending products comes under the domain of "Market Basket Analysis" and "Data Mining". The proposed system will be handled by an e-commerce website admin, wherein the admin will input the name of a customer for whom the e-commerce site wants to recommend products based on similar customer behaviour using data mining.

**Keywords:** Product Recommendation, E-commerce Market Basket Analysis, Data Mining.

## INTRODUCTION

The importance of effective product recommendations cannot be ignored. When done correctly they can heavily contribute to the success of a website, both by increasing the quantity and the size of orders being placed. Studies of personalisation have revealed that, when recommendations are made intelligently, those products being recommended can enjoy conversion rates over 900% higher than site wide averages. Amazon have famously leveraged this technique to attract a huge customer base and vast revenues - a look at their homepage will reveal a site committed to personalising its shopping experience around the needs of each and every individual customer.[1]
Recommendation systems are being used by and increasing number of E-commerce websites. They not only make it easier for consumers to find relevant products to purchase, but also help to engage your shoppers. A personalized product recommendation isn't based on assumption, a guess or random factors and you probably have seen these type of recommendations more than you can imagine - friend suggestions on Facebook, videos on Youtube based on those previously watched, LinkedIn connections based on mutual connections. The examples are endless.
Most of the recommendation systems in e-commerce take either of 3 basic approaches: [2]

**Content-based recommendations:**
In content-based recommendation systems, the buyer expresses some preferences on a set of products and the recommender promotes suggestions based upon a description of the items and profile of user interests. For example, keywords from product description help the system to retrieve information from a retailer's catalogue with the items that share common features. If one user is viewing summer dresses or is likely to leave some comments on this section, content-based filtering can use this history to identify and recommend similar content (other dresses or items that match with summer dresses like shoes).

**Collaborative filtering recommendations:**
Opposite the content-based recommendations, collaborative-filtering algorithms generate recommendations based on other customers who are most similar to that user. The algorithm can measure similarity of multiples customers and make recommendations based upon what certain customers have chosen as relevant. Examples for online retail recommendations would include filters such as "customers who bought this, also bought that" or "customers who looked at this item, also liked this item". In the simplest form, collaborative filtering works best when data from multiple sources like social media, comes together and is sorted into categories. It´s a must for every retailers aiming to personalize their online shopping experience.

**Hybrid recommendations:**
Hybrid algorithms for product recommendations are systems that combine multiple recommendations techniques together to achieve a synergy between them.
Now that recommendation is to be done, finding similarity either on product basis or user basis or both as mentioned above. To calculate similarity various methodologies are used, some of them are:

**Cosine Similarity:**
When products and users are stored in a form of matrix then either products are in a form of vector or users are. To calculate similarity suppose two vectors **A** and **B** the formula for similarity is

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

$$\text{Similarity} = \cos\theta = \frac{\sum_{i=1}^{n} Ai \cdot Bi}{\sqrt{\sum_{i=1}^{n} Ai^2}\sqrt{\sum_{i=1}^{n} Bi^2}}$$

Cosine similarity is one of the most popular similarity measure applied on text documents.[3]

**Jaccard Coefficient:**

This is sometimes referred to as Tanimoto coefficient, formulated as intersection divided by the union. Suppose there are two users **A** and **B**, the similarity coefficient will be calculated as:

$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$

(If A and B are both empty, $J(A,B)=1$)

$$0 \leq J(A,B) \leq 1$$

The Jaccard coefficient is a similarity measure and ranges between 0 and 1. [3]

**Pearson's Correlation Coefficient:**

The Pearson correlation coefficient, often referred to as the Pearson R test, is a statistical formula that measures the strength between variables and relationships. To determine how strong the relationship is between two variables, you need to find the coefficient value, which can range between -1.00 and 1.00.The formula is given as: [3]

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

## METHODOLOGY

In today's world of E-commerce there is increasing need of offering best and extra to the customers who are regular and are actually interested to buy products. This implicates the need for asoftware at the admin end which will automate the data mining for admin and generate result as recommended products which are never bought by that customer, by analyzing the past purchasing behaviour of similar customers.

The system will first take raw data from database (which is .xls file in this system). Based on given transactional data, the system will convert the data into dictionary as shown in below.

```
'riddhesh' : {
                'charger':5,
                'microphone':5,
                'mouse':3,
                'otgcable':2,
                'usbcable':2,
                'scanner':1,
                'printer':5,
                'computer':3,
                'cover':4
            }
```

Fig.1 Python data dictionary

System then reads the username for whom products are to be recommended from GUI and computes similarity with every other user in the dictionary using Pearson Correlation Coefficient and finds the similarity score. The similarity calculated is sorted in descending order and according to this score, recommended products with corresponding similarity score are returned to the GUI application. The back-end functionality is implemented using Python Scripting language. User Interface designed using Tkinter tool. The system will use following algorithms for analysis purposes:-

- User-User Collaborative Filtering: Collaborative Filtering is a very popular Data Mining technique used in recommendation system. This algorithm is used to find the users with similar interest by calculating
- Pearson Correlation Coefficient: This algorithm will calculate similarities between a prospective customer and other customers based on various factors such as: Frequent Purchases, Personal Interests, and Regional Interests etc. Using Pearson Correlation Coefficient gives the advantage that if any user has never bought any product he will be then recommended by the products of an average ranking user's product list. This system gives top 5 results based on the comparison to every user on the list.
- This algorithm automates the import of datasets into the system.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 4, April 2017

Overall System flow:-
1. Loading the CSV dataset: Preprocessed dataset which contains product recommendation data (customer name, list of products purchased, product id, customer, rating)
2. Converting the dataset obtained in the previous step into python supported data dictionary using a python script.
3. Input the dictionary to the actual similarity comparison code in python using Pearson Correlation Coefficient (PCC).
4. Finding PCC to get user with most similar interest, and accordingly finding the products bought by the similar users that can be recommended to the user.
5. Output the results to the GUI.

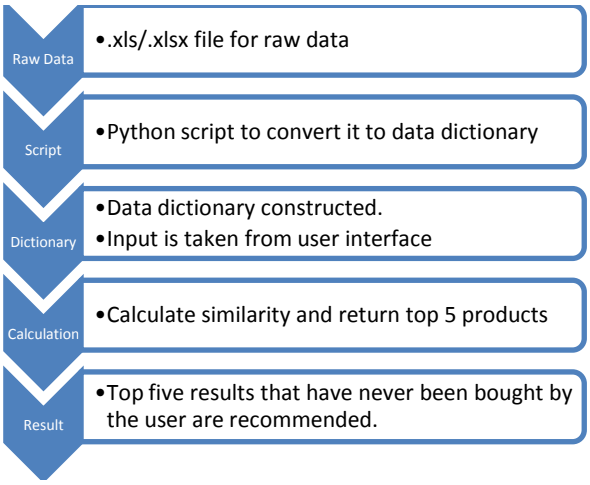


Fig. 2 Overall flow of proposed system.

## RESULTS

The proposed system currently takes .xls/.xlsx format files as raw data set after cleaning is done on data, this format is supported by xlrd/xlwt packages in python. The back-end and front-end part are both implemented using python script. Front end is made using Tkinter package in python. Back end process contains two file, one is for converting .xls/.xlsx files to data dictionary which will be then passed to the second file which will calculate similarity using Pearson Correlation Coefficient between users based on the username (input) given in the user interface. It returns result as names of top five products which are bought by similar users but not by the user.

## CONCLUSION

Studying customer buying patterns makes the business more customer-centric and thus plays a very important role in any e-commerce business. An e-commerce business needs to attract new users and these engines, if improved, can solve this problem so that the business can be more profitable.
The proposed system can further be expanded to different domains like songs, movies etc. Also there are various methods available for finding similarity and recommendation, using the best method the system can recommend most precise result in the context of products, songs, movies etc.

## ACKNOWLEDGMENT

We express our sincere thanks to all the authors, whose papers or algorithms are published in the area of Recommendation Systems, Big data and Data Mining in various conference proceedings and journals. We also express or sincere thanks to all the researchers who are working in the area of Data Mining and Recommendation Systems. And we express our most sincere thanks to the authors whose papers and algorithms have been studied for preparing the review.

## REFERENCES

[1] http://www.smartinsights.com/ecommerce/merchandising/product-recommendations-websites-wrong/ (Accessed on 26/04/17 17:00)
[2] http://insights.strands.com/key-ecommerce-recommendations-algorithms (Accessed on 26/04/17 17:30)
[3] Madhuri Badugu and Bala Krishna,Partitional Clustering using Similarity Measures,International Journal of Computer Trends and Technology-volume-3,Issue-1,92-95,2012