

# **Mining Association Rules for User Recommendation of E-Commerce Websites**

R. Z. InamulHussain<sup>1</sup>, Dr.S. K. Srivatsa<sup>2</sup>, Zaid Sheriff I<sup>3</sup>

Research Scholar, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya University, Enathur, Kanchipuram,  
Tamilnadu, India

Retired Professor, Anna University, Chennai, India

C Abdul Hakeem College Of Engineering and Technology, Melvisharam, Tamilnadu, India

**ABSTRACT:** One of the popular descriptive data mining techniques is Association rule mining owing to its extensive use in marketing and retail communities and many other diverse fields. Mining association rules are particularly useful for discovering relationships among items from large databases. The “market-basket analysis” which performs a study on the habits of customers is the source of motivation behind Association rule mining. The extraction of interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories is the main objective of Association rule mining. With the rapid growth of World Wide Web and electronic commerce, huge volume of data is available on line. The available on line data is plenty and with rich descriptions. Since data are electronically collected they are reliable. So e-commerce is a killer domain for Association rule mining. In the e-commerce environment all the actions of customers visiting a shop from entry to exit are recorded. So customer navigation patterns and their purchasing behaviour are available in the e-commerce data. Finding association rules from these data helps to make right business decisions in right time. It also helps to improve cross selling and web site design. Also Association rule mining in e-commerce data provide navigation and purchasing suggestions to customers. Recommendation system use to generate recommendation of the product that customer may want to buy. This system increase the sale of vendor and easy to find product from available product. In our proposed system we have utilized association rule mining for user Recommendation system in e-commerce environment to buy the required item. In this way it would benefit the sellers to increase their sales and also the customers to get the right product.

## **1.INTRODUCTION**

With the increase the usage of Internet worldwide, the growth of e-commerce business is very high, and buying products from the ecommerce website has become a new trend in modern culture. There is large number of products available in market and it increasing day by day. It is very difficult to find the required product from available large set of product. If customers face difficulties then vendor have risk to lose the customer. In this case, the E-commerce recommendation system becomes very useful for customers as well as vendors. Recommendation system use to generate recommendation of the product that customer may want to buy. This system increase the sale of vendor and easy to find product from available product. Association rule mining use to make real time recommendation system. From the users transaction data-set we can generate rules for customer buying tendency. Based on customer purchased product and customer profile, we can generate recommendation using association rule mining technique. With the development of e-commerce business we want to meet the each and every customer needs individually. Recommendation systems enhance E-commerce sales in three ways as below. A. Converting visitors into purchaser it is general trend that customers are just visit the site and get information about product and forgot it. Recommendation system stores the customers visited product list and give recommendation

whenever customer visit the site. It helps to customers by finding the product which they wish to purchase. B. Increasing Cross-sell Recommender systems use to increase cross-sell of product by recommending products to the customer. After purchasing some product, recommendation system give list of products that may be require for customer. This increases the sale of vendors. C. Building Loyalty Recommendation system use to build relationship between vendor and purchasers. Recommendation system use to learn the customer behavior while purchasing the products. Customers feedback on recommended product is use to for built the loyalty.

## II. MINING ASSOCIATION RULES

There are various kinds of methods to mine the information from the database, such as mining association rules, multi-level data generalization and summarization, classification, and clustering [4]. The most common type is mining association rules.

The problem of mining association rules in databases was first introduced in 1993 by Agrawal [1]. An example of such a rule might be that “90% of customers purchase bread and butter also purchase milk and coffee”. Since its introduction, Association Rules Mining (ARM) [1] has become one of the core data mining tasks. ARM is an undirected or unsupervised data mining technique, which work on massive data, and it produces clear and understandable results. ARM is aimed at finding regularities in data. The following is a formal statement of the problem [1]: Let  $I = \{i_1, i_2, \dots, i_m\}$  be a set of literals, called items. A set of items is also called an itemset. An itemset with  $k$  items is called a  $k$ -itemset. Let  $D$  be a set of transactions, where each transaction  $T$  is a set of items such that  $T \subseteq I$ . Associated with each transaction is a unique identifier, called its  $TID$ . We say that a transaction  $T$  contains  $X$ , a set of some items in  $I$ , if  $X \subseteq T$ . The support of an itemset  $X$ , denoted  $\text{sup}(X, D)$ , is the number of examples in  $D$  where it occurs as a subset. An itemset is *frequent* or *large* if its support is more than a user-specified minimum support ( $\text{min\_sup}$ ) value.

An association rule is an implication of the form  $X \Rightarrow Y$  where  $X \cap Y = \emptyset$  and  $X \cup Y \subseteq I$ .  $X$  is called the antecedent of the rule, and  $Y$  is called the consequence of the rule. The rule  $X \Rightarrow Y$  has support  $s$  in the transaction set  $D$  if  $s\%$  of transactions in  $D$  contains both  $X$  and  $Y$ . That is, the support of the association rules.  $X \Rightarrow Y$  is the probability that  $X \Rightarrow Y$  occurs in the set of transactions in the database  $D$ ; it is denoted by  $\text{support}(X \Rightarrow Y)$ . The rule  $X \Rightarrow Y$  holds in the transaction set  $D$  with confidence  $c$  if  $c\%$  of transactions in  $D$  that contain  $X$  also contain  $Y$ . The confidence of the association rule  $X \Rightarrow Y$  is the probability that a transaction contains  $Y$  given that the transaction contains  $X$ , or it may be given mathematically as  $\text{support}(X \cup Y) / \text{support}(X)$ .

Example 1.1. Consider a set of itemsets  $I = \{A, B, C, D, E, F\}$ . Let  $D$  be a set of four transactions as following:

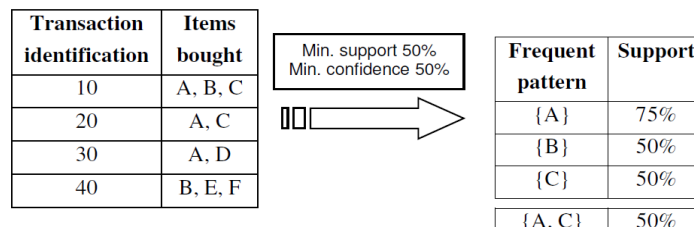


Figure 1. Example database and frequent itemsets

For rule  $A \Rightarrow C$ :

$$\text{support} = \text{support}(\{A\} \cup \{C\}) = 50\%$$

$$\text{confidence} = \text{support}(\{A\} \cup \{C\}) / \text{support}(\{A\}) = 66\%$$

The problem of discovering all association rules can be decomposed into two sub problems:

**International Journal of Innovative Research in Science, Engineering and Technology**

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 7, April 2017

5<sup>th</sup> National Conference on Trends in Automotive Parts Systems and Applications (TAPSA-2017)

30<sup>th</sup> & 31<sup>st</sup> March 2017

Organized by

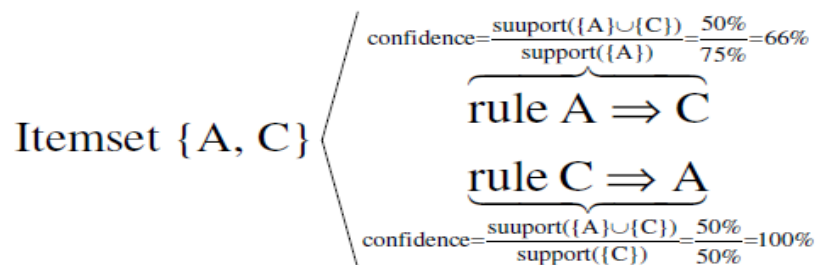
Sri Krishna College of Engineering & Technology, Kuniamuthur, Coimbatore-641008, Tamilnadu, India

1. Find all acts of items (*itemsets*) that have transaction support above minimum support. The support for an item is the number of transactions that contain the itemset. Recall that an itemset is *frequent* or *large* if its support is more than a user specified *minimum support* (*min\_sup*) value.

*Example 1.2.* From the above database, we obtain four frequent itemsets {A}, {B}, {C} and {A, C} with supports of 75%, 50%, 50% and 50% respectively.

2. Use the *large* itemsets to generate the desired rules. Here is a straightforward algorithm for this task. For every large itemset *l*, find all non-empty subsets of *l*. For every such subset *a*, output a rule of the form  $a \Rightarrow (l - a)$  if the ratio of *support*(*l*) to *support*(*a*) is at least *minconf*. We need to consider subsets of *l* to generate rules with multiple consequents.

*Example 1.3.* From the frequent itemset {A, C} found in example 1.2, we can generate two rules whose confidences are greater than or equal to *minconf* value.



**III. MINING ASSOCIATION RULES WITH WEIGHTED ITEMS**

In the last section, we discussed about mining association rule for unweighted case. In the following, we introduce the conceptual framework of weight and apply it to mining association rules.

There have been two approaches to the association rule mining problem. The first one is to mine the frequent itemsets regardless of their coefficients [1, 7]. The second trend is to assign weights to the items to reflect their importance to the users. Some previous works focused on mining frequent itemsets with weighted items [5] and different supports [6].

*Example 2.1.* The following rule is the unweighted binary association rule from [1]:

$$(\text{Bread} = \text{Yes}) \Rightarrow (\text{Ham} = \text{Yes})$$

with support = 60% & confidence = 80%

The above rule states that the probability of buying bread and ham in a set of transaction is 0.6, and the *confidence* states that probability that buying ham, given that customer buys bread, is 0.8.

The above rule is an unweighted case. However, it is better for the following cases to consider the importance of the items or attributes.

For example, the rule

$$(\text{Income} = \text{High}) \Rightarrow (\text{School level} = \text{High})$$

is, in human interpretation, probably more interesting than

**International Journal of Innovative Research in Science, Engineering and Technology**

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 7, April 2017

5<sup>th</sup> National Conference on Trends in Automotive Parts Systems and Applications (TAPSA-2017)

30<sup>th</sup> & 31<sup>st</sup> March 2017

Organized by

Sri Krishna College of Engineering & Technology, Kuniamuthur, Coimbatore-641008, Tamilnadu, India

$$(Price = High) \Rightarrow (Sales = Low)$$

Even if the support of the latter rule is much more than that of the former. By using the weights, the importance of the attributes or items can be reflected, and we can mine the rules with interestingness. For example, we can add the weights to the sales transactions, where the items are under promotion, or with more profits. The unweighted association rules would be the same if the database did not change, thus it cannot provide a flexible way for the users to adjust the priority of the rules. Therefore, the mining association rules for weighted items were presented in [6] to resolve this problem.

TID	Product ID	TID	Product ID
1	1 4	2	1 4 5
3	2 3 5	4	3 5
5	1 2 4 5	6	1 3 4
7	3	8	2 5

**Table 2.** Transactions of a stationery store

*Example 2.2.* Suppose in a stationery store, a database is shown in Table 1. Each item includes information of name, profit and given weight. Table 2 gives the transaction database. For each transaction, there will be a transaction identifier (*TID*) and the names of items. Suppose there are only 5 items and totally 8 transactions in the transaction database.

Regardless of the weights given, if the value of *minsup* is set to 0.4, {1, 4} will be large itemset since its support is 50%. However, {1, 4, 5} is not a large itemset as it appears only two times in the database.

But if we take weights of items into account, and the given value of *wminsup* is 0.4, {1, 4} will not be a large weighted itemset since

$$(0.1 + 0.5) \times \frac{4}{8} = 0.3 \leq 0.4$$

On the contrary, {1, 4, 5} will be a large itemset since

$$(0.1 + 0.5 + 1) \times \frac{2}{8} = 0.4 \geq 0.4$$

By the same argument, {5}, {1, 2, 5} will be large weighted itemsets. Although itemset {1, 4} has a greater support than that of {1, 2, 5}, it seem to be true that the latter otherwise make a greater profit than the former can do. In this case, we say that itemset {1, 2, 5} is more interesting than itemset {1, 4}, or the *interestingness* of itemset {1, 2, 5} is greater than that of itemset {1, 4}.

#### IV. MINING ASSOCIATION RULES WITH ADJUSTABLE

##### INTERESTINGNESS

##### 4.1 Interestingness and interesting itemsets

Based on the definitions of weighted itemsets in previous chapter, we extend the definitions of interestingness and interesting itemsets.

**International Journal of Innovative Research in Science, Engineering and Technology**

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 7, April 2017

5<sup>th</sup> National Conference on Trends in Automotive Parts Systems and Applications (TAPSA-2017)

30<sup>th</sup> & 31<sup>st</sup> March 2017

Organized by

Sri Krishna College of Engineering & Technology, Kuniamuthur, Coimbatore-641008, Tamilnadu, India

**Definition 1.** The interestingness of an itemset  $X$ , denoted  $interest(X)$ , is the coefficient correlation between the number of transactions in which it occurs as a subset and the total weight of its items, or mathematically,

$$interest(X) = \left( \sum_{j \in X} w_j \right) \cdot support(X)$$

In order to find the interesting itemsets, the threshold, minimum interestingness ( $min\_int$ ) must be specified.

**Definition 2.** An itemset  $X$  is called an interesting itemset if the interestingness of the itemset  $X$  is greater than or equal to the interestingness threshold, or mathematically,

$$interest(X) \geq min\_int$$

*Example 3.1.* From the database in Table 1 and 2, we can calculate the interestingness of itemsets as the following table. The itemsets are sorted into descending order of their interestingness.

Itemset	W*	S*	I* ↓
{5}	1	62.5%	0.625
{2, 5}	1.2	37.5%	0.45
{1, 4, 5}	1.6	25%	0.4
{4, 5}	1.5	25%	0.375
{3, 5}	1.3	25%	0.325
{1, 4}	0.6	50%	0.3
{1, 5}	1.1	25%	0.275
{4}	0.5	50%	0.25
{2, 4, 5}	1.7	12.5%	0.2125
{2, 3, 5}	1.5	12.5%	0.1875
{1, 2, 5}	1.3	12.5%	0.1625
{1, 3, 4}	0.9	12.5%	0.1125
{1, 2, 4}	0.8	12.5%	0.1
{3, 4}	0.8	12.5%	0.1
{2, 4}	0.7	12.5%	0.0875
{2}	0.2	37.5%	0.075
{2, 3}	0.5	12.5%	0.0625
{1}	0.1	50%	0.05
{1, 3}	0.4	12.5%	0.05
{1, 2}	0.3	12.5%	0.0375
{3}	0.3	50%	0.15

\* W, S and I are acronyms for Weight, Support and Interestingness, respectively.

**Table 3.** Itemsets sorted into descending order of their interestingness

If the value of  $min\_int$  is 0.3, we obtain six interesting itemsets; these are: {5}, {2, 5}, {1, 4, 5}, {4, 5}, {3, 5}, {1, 4}. Of these six interesting itemsets, five contain item 5 which represents for pens. It proves that the interestingness of an itemset is made up of its weight and support.

**V. RELATED WORK**

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve change over and evaluation of changeover methods.

## International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 7, April 2017

5<sup>th</sup> National Conference on Trends in Automotive Parts Systems and Applications (TAPSA-2017)

30<sup>th</sup> & 31<sup>st</sup> March 2017

Organized by

Sri Krishna College of Engineering & Technology, Kuniamuthur, Coimbatore-641008, Tamilnadu, India

### MODULE DESCRIPTION

#### GENERATION OF ASSOCIATION RULES

- Association rules are generated and it is used for product recommendation.
- The frequently viewed and bought items are found and it used to recommend people to buy.
- Usage of association rules in ecommerce enhances the business market.
- It would benefit the sellers to increase their sales and also the customers to get the right product.

#### BASIC TERMINOLOGY AND MATHEMATICS

##### Association Rules

- A series of methodologies for discovering interesting relationships between a variables in a database
- The outcome of this technique, in simple terms is a set of rules that can be understood as “**if this, then that**”.
- An example of association rule would be:
- If a person buys peanut butter and bread then he will also purchase jelly.
- A rule is typically written in the following format:  $\{ i_1, i_2 \} \Rightarrow \{ i_k \}$
- The  $\{ i_1, i_2 \}$  represents the left hand side, LHS, of the rule and the  $\{ i_k \}$  represents the right hand side, RHS. This statement can be read as “if a user buys an item in the item set on the left hand side, then the user will likely buy the item on the right hand side too”.
- A more human readable example is:  
 $\{ \text{coffee, sugar} \} \Rightarrow \{ \text{milk} \}$
- If a customer buys coffee and sugar, then they are also likely to buy milk.

##### Support

- Support is an indication of how frequently the itemset appears in the database.
- Ex. A grocer has 10 transactions in total. Of which  $\{ \text{Bread, Butter} \}$  appears 4 times. The support for this rule is  $4/10$  or  $0.4$ . It has a support of 40%.

##### Confidence

- Confidence is an indication of how often the rule has been found to be true.
- $\text{Confidence}(X \rightarrow Y) = \text{Support}(X \square Y) / \text{Support}(X)$
- $\text{Confidence}(\text{Bread, Butter} \Rightarrow \text{Milk}) = 0.4 / 0.4 = 1.0$
- This means that 100% of the transactions containing butter and bread the rule is correct (100% of the times a customer buys butter and bread, milk is bought as well).

##### Lift

- $\text{Lift}(X \rightarrow Y) = \text{Support}(X \square Y) / (\text{Support}(Y) * \text{Support}(X))$
- $\text{Lift}(\text{Milk}, \text{Bread} \Rightarrow \text{Butter}) = 0.2 / (0.4 * 0.4) = 1.25$

Fig. 6.1 shows the results obtained on retail (.gz). raw dataset, using numerous values of support ranging from 0.05 to 0.2. It can be obviously observed that modified and proposed algorithm produces much better reduced and required rulesthan FP-Growth, depending on the support used.

Support	FP-Growth	Modified
0.05	27	22
0.075	21	17
0.1	10	7
0.125	9	8
0.15	8	6
0.175	7	5
0.2	5	4

Table 6.1 Generated Rules

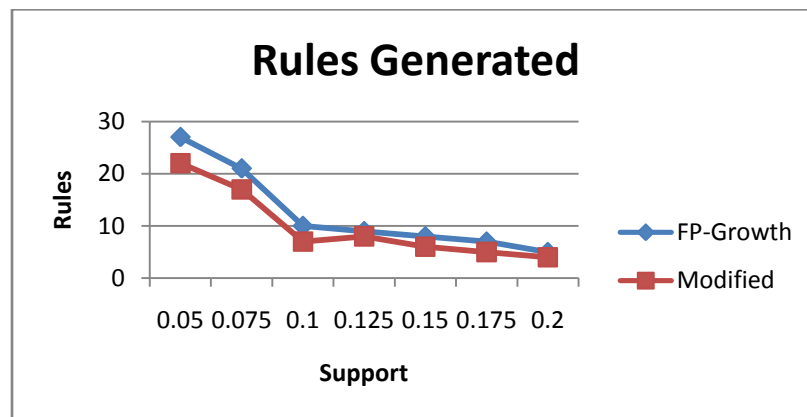


Figure 6.1 Number of Rules Generated

### CREATION OF E-COMMERCE WEBSITE

WooCommerce is a free ecommerce plugin that allows you to sell anything, beautifully. Built to integrate seamlessly with WordPress, WooCommerce is the world’s favouriteecommerce solution that gives both store owners and developers complete control.

### SELL ANYTHING, ANYWHERE

With WooCommerce, you can sell both physical and digital goods in all shapes and sizes, offer product variations, multiple configurations, and instant download to shoppers, and even sell affiliate goods from online marketplaces.

## International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 7, April 2017

5<sup>th</sup> National Conference on Trends in Automotive Parts Systems and Applications (TAPSA-2017)

30<sup>th</sup> & 31<sup>st</sup> March 2017

Organized by

Sri Krishna College of Engineering & Technology, Kuniyamuthur, Coimbatore-641008, Tamilnadu, India

With premium extensions, you can offer bookings, memberships, and recurring subscriptions. Perhaps you'd like to sell monthly subscriptions for physical goods, or offer your members a discount on digital downloads? It's all possible.

### SHIP WHEREVER YOU LIKE

Offer free shipping, flat rate shipping, or make real-time calculations. Limit your shipments to specific countries, or open your store up to the world. Shipping is highly configurable, and WooCommerce even supports drop shipping.

### EXTENSIVE PAYMENT OPTIONS

WooCommerce comes bundled with the ability to accept major credit cards, PayPal, BACS (bank transfers), and cash on delivery. Need additional options? More than 140 region-specific gateways integrate with WooCommerce, including popular choices like Stripe, Authorize.Net, and Amazon Payments.

### USER RECOMMENDATION ENGINE

We use woocommerce recommendation engine that utilizes association rules for user recommendation. WooCommerce Recommendation Engine automatically recommends products to users based on view history, purchase history and products that are frequently purchased together.

Your visitor will enjoy recommendations that are formed by other visitors interacting your web site. Your visitors are unique. They have their own tastes, needs and backgrounds. It is aimed at providing personalized feeling when they visit your site. It allows your very own customers help each other to discover the best product in your websites for themselves. The recommendation is served automatically according to the visitors actions (Collaborative Filtering). Those actions are viewing product and buying the products.

For example:

Visitor A is viewing the product detail page of product 1, product 2, and product 3. This visitor A actions will make those three product to have products recommendation. For other visitors who visit your site, when they view the product detail age for product 1, product 2, or product 3. The recommendation based on the previous visitor, visitor A, will be displayed. In this case, if detail page of product 1 is viewed, the recommendation contains product 2 and product 3. When product 2 detail page is viewed, the recommendation contains product 1 and product 3. When the detail age of product 3 is viewed the recommendation contains product 1 and product 3.

Fig. 6.2 shows the product recommendation given to the user. This system automatically recommends products to the user based on purchase history and products that are frequently purchased together. It is a great way to provide for automatic cross and up sells, and will help users browse and purchase more products.



## International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization

Volume 6, Special Issue 7, April 2017

5<sup>th</sup> National Conference on Trends in Automotive Parts Systems and Applications (TAPSA-2017)

30<sup>th</sup> & 31<sup>st</sup> March 2017

Organized by

Sri Krishna College of Engineering & Technology, Kuniamuthur, Coimbatore-641008, Tamilnadu, India

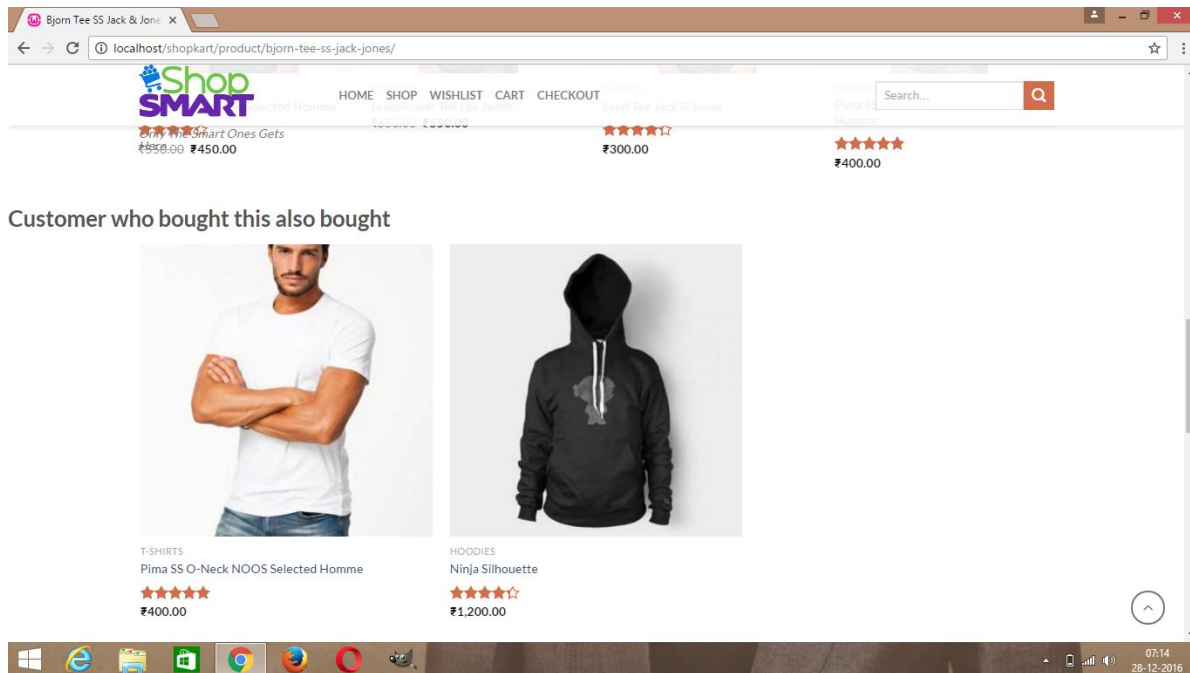


Figure 6.2 Recommendation given to the user

## VI. FUTURE ENHANCEMENT

### AI Systems

Our future work will be the application of AI systems, running evolutionary algorithms designed to find the absolute optimum. By developing systems using these algorithms, which are designed to effectively test, optimise and repeat on loop, ecommerce will move into an even higher level of sophistication as this technology improves.

### Drone Delivery

Companies have been working their way around to innovate the delivery process to shorten human effort as well as time. The answer to these problems is Delivery by Drones. DGCA is now fast tracking the process of issuing guidelines for the use of drones for civil purposes in India. If everything goes as per the plan, then India might become the first country in the world to allow the use of drones for civil purposes.

### App only Approach

Statistics suggest the future of internet lies in mobiles. Experts say more than 580 million people in India will use the Internet by 2018, and 70-80% of them will access the Web on mobile phones.

## REFERENCES

- [1] R. Agrawal, T. Imielinski, and A. Swami, 'Mining association rules between sets of items in large databases'. In Proc. of the ACM SIGMOD Conference Management of Data, Washington D.C., May 1993.
- [2] P. Adriaans, D. Zantinge, 'Data mining', Addison-Wesley, 1999.

**International Journal of Innovative Research in Science, Engineering and Technology**

*An ISO 3297: 2007 Certified Organization*

*Volume 6, Special Issue 7, April 2017*

**5<sup>th</sup> National Conference on Trends in Automotive Parts Systems and Applications (TAPSA-2017)**

**30<sup>th</sup> & 31<sup>st</sup> March 2017**

**Organized by**

**Sri Krishna College of Engineering & Technology, Kuniamuthur, Coimbatore-641008, Tamilnadu, India**

- [3] J. Han, M. Kamber, 'Data Mining: Concepts and Technique', University of Illinois, 2002
- [4] U. Fayyad, S. Chaudhuri, P. Bradley, 'Data mining and its role in databasesystems', 1999
- [5] D. V. Thanh, P. T. Hoan, P. X. Hieu, N. T. Trung[Mining association rules with different supports], Conference of junior scientists of Vietnam Nat'l Univ. Hanoi, pages475-483,2002
- [6] C. H. Cai, 'Mining association rules with weighted items', Thesis for degree of master, Chinese University of Hong Kong, 1998
- [7] <http://www.woocommerce.com>
- [8] <http://wordpress.org>
- [9] <http://www.wordpress.org/plugins>
- [10] <https://woocommerce.wordpress.com/2015/08/07/api-settings-and-the-api>