

Customisable Bundling Approach for Online Supermarkets using Association Rules of Product Categories

Vani V Nair^{1*}, Vedashree V², Vimala B K³, Yamuna M⁴, ChetanaSrinivas⁵

^{1,2,3,4,5}Student, Department of Computer Science, East West Institute of Technology, Bengaluru, India

DOI: <https://doi.org/10.26438/ijcse/v7si15.138143> | Available online at: www.ijcseonline.org

Abstract—This research deals with the identification of customers and their buying behavior patterns. The aim is to sell the products which are least preferred by the customers so as to make a cost-effective sale by using bundling approach. A Customized bundling is a group of resources joined together in a single package that has an associated logical name. A bundle is a collection of products which are sold together for a single price. It is the well organized way to make the customer's shopping self-satisfied. It is implemented by the integration of associative clustering and Support vector machine (SVM) with java.

Keywords-Bundling, Associative clustering, Support vector machine, Suggestions.

I. INTRODUCTION

The market now a days is getting very competitive. Organizations are spending lot of money in planning to sell their products in the most effective way[1]. For many years vendors had been operating at low profit. Online supermarkets are facing loss due to higher operational cost required for delivery, order picking and packing. Therefore, how to increase the profit and reduce operating cost has become a major challenge for online supermarkets. This can be solved using Data mining techniques with the bundling approach. Data mining is a process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends. By “mining” large amount of data, hidden information can be discovered and used for other purposes. Online shopping is a form of electronic commerce which allows consumers to directly buy goods or services from a seller over the internet using a web browser. Consumers find a product of interest by visiting the website of the retailer directly or by searching among alternative vendors using a shopping search engine. Bundling approach is a concept of tying products together to create an appealing package for customers according to their need. This provide a services to a customer, from which customer derives better value than if the product were purchased individually.

Customized bundle products is a marketing strategy where a seller combines multiple products and offer the bundled items as one product, in which customers can customize product variants and quantity in a bundle according to need. For example, when a customer selects a product, the

products which are frequently bought together are suggested as a bundle. Within the bundle, the customer is able to view more than one similar products of different brands, prices and other features. According to the customers' needs, he or she can add the product to the bundle and overall price will be given. This helps the customer find the product which they want along with the ones they might need at a better price. The vendor can try to sell the products alongside the ones which are sold which thereby getting profit and also by reducing shipping charges of individual products.

The rest of the paper is organized as follows. In Section 2 literature survey is done. In Section 3, proposed method is described. In Section 4, experimental results are shown. Finally in Section 5, the conclusion of our work is discussed.

II. RELATED WORK

Association rule mining and Apriori algorithm is applied for the customer behavior analysis. Support and confidence is the result of the Apriori algorithm which is implemented on the association of different product, like $A \rightarrow B$ and $B \rightarrow A$, support and confidence is gained from this association[1]. Yang and Lai [3] have investigated the method of product bundling for an online book store with different data sources. They pointed out that by integrating the customer orders, browsing data and data of shopping-cart, more interesting product bundles can be provided to online consumers. A cluster algorithm for exterminating the customers' lifestyles has been proposed by Miguéis et al. [4]. Then each customer was classified into different lifestyle cluster according to the rule of closet similarity of the customer past purchase history. The algorithms of K-means and association rules are

proposed by Liao et al. [5] as a bundling solution for marketing.

In conclusion, the existing data mining methods provide us with popular product combinations and consumption patterns[2]. However, due to the large amount of SKUs in the online supermarkets, and various preferences of a large number of customers, it also suffers from a very large number of unnecessary frequent itemsets or consumption patterns. Our approach integrate two aspects:

(1) process to have combination of high sold product and low sold product to get profit.

(2) to extract or fetch the data from database easily.

III. METHODOLOGY

Based on association rules and recommendation technique, the method to recommend a customized bundle is proposed. The basic principles of the method is to combine association rules of product categories and personal recommendation techniques. The association rules are used to generate a serious correlated product categories for customers. For example, in figure 1, three product categories are provided for the consumer, which is based on association rules of product categories. Personal recommendation techniques is used to offer a number of products in each category instead of too many products. For example, in figure 1, there are three products available in each category. The procedures are shown in figure 2 and explained as follows.

A. User Module

First create login module and in this login frame module existing user login with their username and password. Suppose if user does not exist then click on register button that will move to Register Module if does not exist then create register module. In Register Module all fields are mentioned like first name, middle name, last name, email, password, phone number, user name etc.

B. Data Extraction

Data Extraction is a act or process of retrieving a data out of sources for further data processing or data storage. In this step we are reading the dataset by using java POI API fetching all record from the dataset.

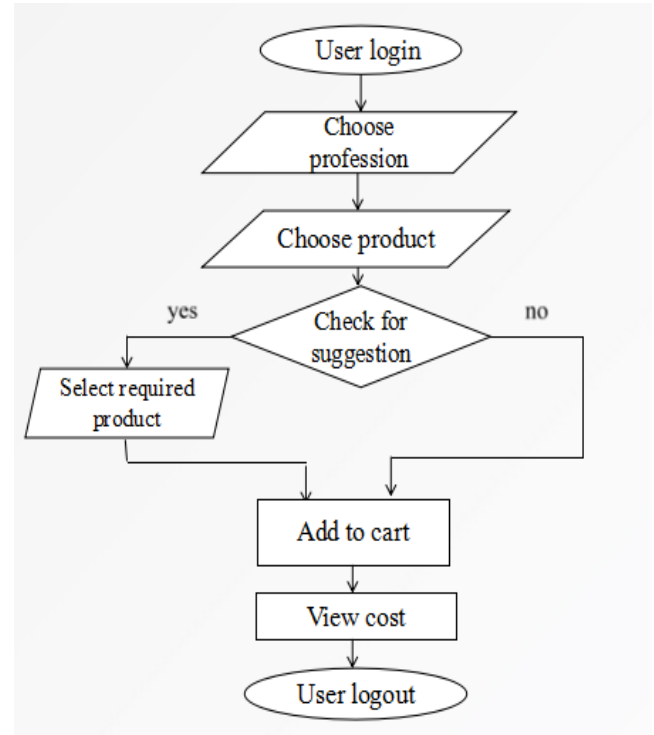


Figure1 : The proposed method

To read or write an Excel, Apache provides a very famous library POI. This library is capable enough to read and write both XLS and XLSX file format of Excel. Excel file can be read by Java IO operation. For that, we need to use Apache POI Jar. POI has different Interfaces Workbook, Sheet, Row, Cell.

C. Data Clustering

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters. In this step we are taking all country information .and grouping all Country into cluster.

D. Associative clustering

Clustering is the task of dividing the population or data points into number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In associative clustering we cluster two sets of data , with samples x and y separately such that clustering would capture as much as possible of dependencies within data pairs(x,y) and the clusters would contain similar data points. Associative clustering is a method for separately clustering two data sets when one-to-one associations between the sets, implying statistical

dependency are available. The pseudocode for the algorithm Associative clustering is given in algorithm 1.

Algorithm 1 Associative clustering

INPUT : $\sum NA$ // Association

OUTPUT : $\{C$

```

1 : N<=0    //Size of node1
2 :  $\{SC$  Items <=0
3: N<=Size(NA)
4 :     For i in 1:n
5 :         Start
6 :         Items=Fetch(NA(i)) //Fetching the items
7 : P1 = Item(i)
8 : P2 = Item (j)
9 :  $\sum L1$  = RETRIEVE [P1]
10 :  $\sum L2$  = RETRIEVE [P2]
11 : SC Item[] = Association [L1,L2]
12 : SC <= SC Items
13 : End for
14 : FRAME(SC)

```

The input is the set of all data points, in this case all the customer data which is retrieved from the supermarket and stored in excel sheets. The algorithm initializes the size of node to less than or equal to 0. All the items are fetched and kept in an arraylist. Clusters are created by grouping similar products together. The association of the different data points within each cluster is studied. This process is repeated until the last item in the excel sheet is processed.

E. Support Vector Machine

Support Vector Machines (SVMs) are supervised learning methods used for classification and regression tasks that originated from statistical learning theory. A classification task usually involves training and test sets which consist of data instances. Each instance in the training set contains one target value (class label) and several attributes (features). The goal of a classifier is to produce a model able to predict target values of data instances in the testing set, for which only the attributes are known. The below picture is an example of a situation in which various linear classifiers can separate the data. However, only one maximizes the distance between itself and the nearest example of each class (i.e. the margin) and for that is called the optimal separating hyperplane. It is intuitively expected that this classifier generalizes better than the other options. The basic idea of SVM classifier uses this approach, i.e. to choose the hyperplane that has the maximum margin.

Algorithm

Let D be a classification dataset with n points in a d -dimensional space $D = \{(x_i, y_i)\}$, with $i = 1, 2, \dots, n$ and let there be only two class labels such that y_i is either $+1$ or -1 .

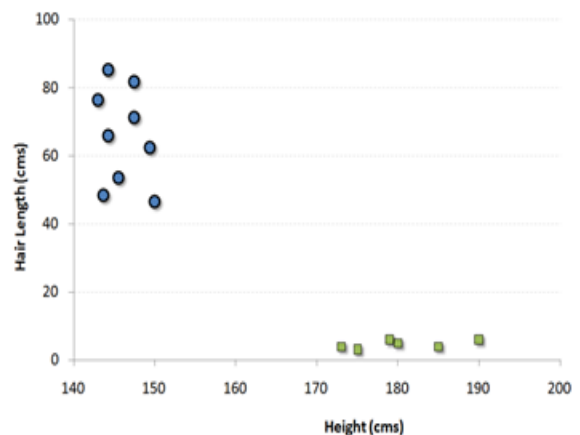


Figure2 :Test set

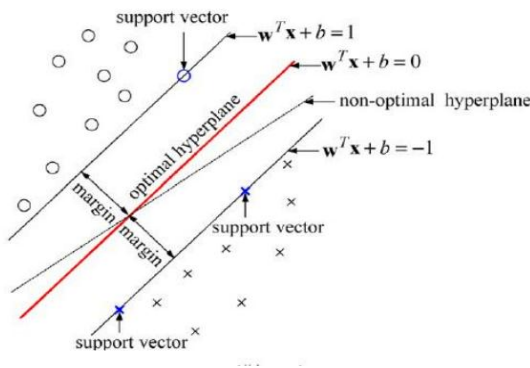


Figure 3:Optimal separating hyperplane

A hyperplane $h(x)$ gives a linear discriminant function in d dimensions and splits the original space into two half-spaces:

$$h(x) = w^T x + b = w_1 x_1 + w_2 x_2 + \dots + w_b x_b \tag{1}$$

where w is a d -dimensional weight vector and b is a scalar bias. Points on the hyperplane have $h(x) = 0$, i.e. the hyperplane is defined by all points for which $w^T x = -b$.

if the dataset is linearly separable, a separating hyperplane can be found such that for all points with label -1 , $h(x) < 0$ and for all points labeled $+1$, $h(x) > 0$.

Given a separating hyperplane $h(x) = 0$, it is possible to calculate the distance between each point x_i and the hyperplane by:

$$\Delta_i = \frac{y_i (h(x_i))}{\|w\|} \tag{2}$$

The margin of the linear classifier is defined as the minimum distance of all n points to the separating hyperplane.

$$\Delta^k = \min_{x_i} \left\{ \frac{y_i (h(x_i))}{\|w\|} \right\} \tag{3}$$

IV. RESULT AND DISCUSSION

In this section, all the steps in each phase of the research methodology are investigated in detail. The company aims to recommend customized product bundles to its customers.

The proposed system is tested on the datasets containing transactions of the company with its customers from a local online supermarkets. This research was carried out taking the information of 456 customers and their transactions with the supermarket. The obtained information was stored in excel sheets and retrieved using Java Apache POI API library files. To read or write an Excel, Apache provides a very famous library POI. Table 1 shows the example of data stored in the excel sheet.

Clustering of similar products is done using associative clustering. One of the machine learning concept known as SVM is used for classification and regression problems. It is used in this project for giving related suggestions and bundles according to the input of customer. For training, 200 customer information were used. For testing, the proposed method 256 customer information were used.

From the experimental results we came to know that the user can successfully login into the home using login id and

password or by getting registered. Once login to the home page, user is now able to select the profession according to which he or she can select the required products. After selecting the products the user can fetch the suggestions from bundle suggestion where he or she can see number of bundle items with different brands and prices. If the user needed any of those bundle, can add to the cart and check for the total cost and he or she will be able to successfully logout.

TABLE 1: EXAMPLE OF TRANSACTION RECORDS

Quantity	Price	Customer id	Gender	Job	Category	Age
5	2.55	1123	male	doctor	General medicine	35
2	1.8	1124	female	engineer	Tools	24
6	3.2	1125	female	professor	Books	28

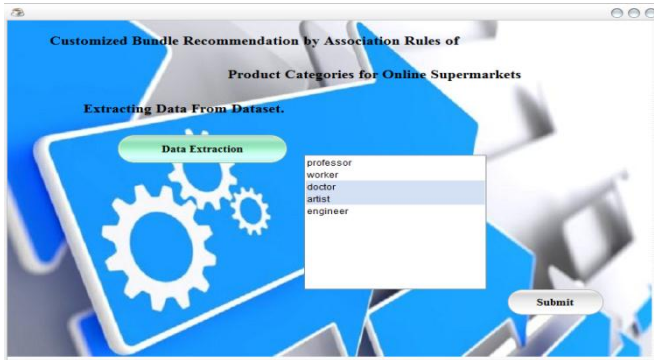


Figure 4: Data Extraction

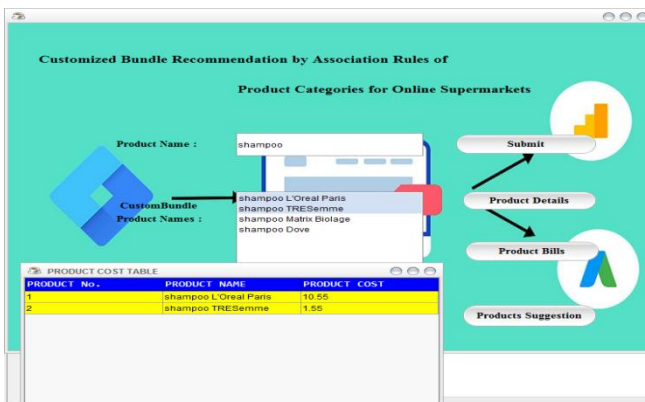


Figure 5: Data Collection



Figure 6: Data Clustering

V. CONCLUSION AND FUTURE WORKS

In this paper, a bundling technique for online supermarkets has been proposed. The customer is able to get suggestions on products which are frequently bought together and also customize the bundle. The time spent by customer on choosing products and also the overall operational cost is reduced. This paper is based on extraction technique and associative clustering. Unwanted products can be reduced using associative clustering. This method is profitable for both vendor and buyer.

The future work includes the implementation of the system in android which will help improve the availability of the system to more vendors.

ACKNOWLEDGMENT

Firstly, we express our sincere thanks to our guide Mrs.ChetanaSrinivas, Assoc Professor, Department of CSE, EWIT and Dr.ArunBiradar, Head of Department, Computer science and engineering for their moral support. We express our sincere gratitude to our principal Dr. K Chennakeshavalu for his constant support and encouragement, we also thank all the faculties of East West Institute of Technology for their co-operation and support.

REFERENCES

- [1]Customer behavior analysis towards online hopping using data mining.,Muqaddas Gull and Arshi Pervaiz,©2018 IEEE.
- [2] Customized Bundle Recommendation by Association Rules of Product Categories for Online Supermarkets.,Yan Fang,Xinyue Xiao,Xiaoyu Wang,Huiqing Lan,©2018.
- [3] Yang, T. C., & Lai, H. (2006). Comparison of product bundling strategies on different online shopping behaviors. *Electronic Commerce Research and Applications*, 5(4), 295–304.
- [4] Miguéis, V. L., Camanho, A. S., &Falcão e Cunha, J. (2012). Customer data mining for lifestyle segmentation. *Expert Systems with Applications*, 39(10), 9359–9366.
- [5] Liao, S. H., Chen, Y. J., & Hsieh, H. H. (2011). Mining customer knowledge for direct selling and marketing.*Expert Systems with Applications*, 38(5), 6059–6069.
- [6]Fuloria , S . (2011) How Advanced Analytics Will Inform and Transform U.S. Retail. *CognizantReports*, July , <http://www.cognizant.com/InsightsWhitepapers/How-Advanced-Analytics-Will-Inform-and-Transform-US-Retail.pdf> , accessed January 2012 .
- [7] Bain Consulting. *China Shoppers Report 2017*. Beijing: Bein Consulting, 2017.
- [8] Jasin S, Sinha A. An LP-based correlatedrounding scheme for multi-item ecommerce orderfulfillment[J]. *Operations Research*, 2015, 63(6): 1336-1351.
- [9] Karimi-Majd, A.-M., &Fathian, M. (2017). Extracting new ideas from the behavior of social network users. *Decision Science Letters*, 6(3), 207–220.
- [10] Miguéis, V. L., Camanho, A. S., &Falcão e Cunha, J. (2012). Customer data mining for lifestyle segmentation. *Expert Systems*

with Applications, 39(10), 9359–9366.

- [11] Liao, S. H., Chen, Y. J., & Lin, Y. T. (2011). Mining customer knowledge to implement online shopping and home delivery for hypermarkets. *Expert Systems with Applications*, 38(4), 3982–3991.
- [12] Abouee-Mehrzi H, Berman O, Sharma S (2014) Optimal joint replenishment and proactive transshipment policies in a multiperiod inventory system. Working paper, University of Waterloo, Waterloo, Ontario.
- [13] Adelman D (2003) Price-directed replenishment of subsets: Methodology and its application to inventory routing. *Manufacturing Service Oper. Management* 5(4):348–371.
- [14] Adelman D, Nemhauser GL, Padron M, Stubbs R, Pandit R (1999) Allocating fibers in cable manufacturing. *Manufacturing Service Oper. Management* 1(1):21–35.
- [15] Agatz N, Fleischmann M, van Nunen J (2008) E-fulfillment and multi-channel distribution—A review. *Eur. J. Oper. Res.* 187(2): 339–356. [16] Alptekinoglu A, Tang C (2005) A model for analyzing multichannel distribution systems. *Eur. J. Oper. Res.* 163(3):802–824.
- [17] Amazon.com, Inc. (2012) Form 10-K for 2011. Filing date: February 1, 2012. U.S. Securities and Exchange Commission, Washington, DC.
- [18] Cattani K, Gilland W, Heese H, Swaminathan J (2006) Boiling frogs: Pricing strategies for a manufacturer adding a direct channel that competes with the traditional channel. *Production Oper. Management* 15(1):40–56.
- [19] Forrester Research, Inc. (2014b) European online retail forecast: 2013 to 2018. Report, May 29. Forrester Research, Inc., Cambridge, MA.
- [20] Gans N, van Ryzin G (1999) Dynamic vehicle dispatching: Optimal heavy traffic performance and practical insights. *Oper. Res.* 47(5):675–692.
- [21] Kaplan A (1969) Stock rationing. *Management Sci.* 15(5):260–267.
- [22] Lee H (1987) A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Sci.* 33(10):1302–1316.

Authors Profile

Ms. Vani V Nair is pursuing her 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. Her area of interest includes Data Mining.

Ms. Vedashree V is pursuing her 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. Her area of interest includes Data Mining.

Ms. Vimala B K is pursuing her 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. Her area of interest includes Data Mining.

Ms. Yamuna M is pursuing her 8 semester B.E in Computer Science & Engineering at East West Institute of Technology, Bengaluru, India. Her area of interest includes Data Mining.

Mrs. Chetana Srinivas got M.Tech degree in Computer Science, Bengaluru, India. She is currently pursuing phd in vtu university and working as Associate Professor in the Department of CSE, EWIT. Her area of interest includes Image Processing, Data Mining, Big Data.