ELSEVIER

A Machine Learning Framework for Predicting Purchase by online customers based on Dynamic Pricing

Rajan Gupta* and Chaitanya Pathak**

*Department of Computer Science, Faculty of Mathematical Science, University of Delhi, Delhi – 110007, INDIA
**Head of Products, Ask-me-Bazaar Online Marketplace, Delhi, INDIA

Abstract

Pricing in the online world is highly transparent & can be a primary driver for online purchase. While dynamic pricing is not new & used by many to increase sales and margins, its benefit to online retailers is immense. The proposed study is a result of on-going project that aims to develop a generic framework and applicable techniques by applying sound machine learning algorithms to enhance right price purchase (not cheapest price) by customers on e-commerce platform. This study focuses more on inventory led e-commerce companies, however the model can be extended to online marketplaces without inventories. Facilitated by statistical and machine learning models the study seeks to predict the purchase decisions based on adaptive or dynamic pricing of a product. Different data sources which capture visit attributes, visitor attributes, purchase history, web data, and context understanding, lays a strong foundation to this framework. The study focuses on customer segments for predicting purchase rather than on individual buyers. Personalization of adaptive pricing and purchase prediction will be the next logical extension of the study once the results for this are presented. Web mining and use of big data technologies along with machine learning algorithms make up the solution landscape for the study.

1. Introduction

Dynamic pricing or price optimization is the concept of offering goods at different prices which varies according to the customer's demand. The pricing of the commodity can be done on the basis of competitor's pricing, supply, demand and conversion rates and sales goals [1]. The art of dynamic pricing is sometimes also referred to as individual level price discrimination [2], revenue management [3] and yield management. Further, adjustment of the prices attributed towards customer's will [4] is another way of defining dynamic pricing. Additionally, customization of the inventory goods by segmenting the customers on the basis of product choice and thus proffering different prices to them is dynamic pricing [5]. It is also referred as real time pricing, wherein value of a product is determined by the current market conditions under commercial transactions. It is a blanket term for shopping experience which defines the prices of the products according to the competitive environment among the suppliers, time of the day and the weather conditions [6].

Dynamic pricing is a wide spread phenomenon whose influence can be seen in industries like retail, automotive, mobile communication, electricity, air ticket and a lot many. The rise in the retail industry [7] has been due to the increased availability of demand data of the customers, the new technologies helping determine prices more

efficiently by studying the consumer pattern and the decision support tools due to the new emerging technology.

The influence, experienced in the mobile communication sector [8], can be attributed towards the decreased call rates, higher competition level and improved network infrastructure. Moreover, the impact felt in the automotive industry is the result of enhanced coordination among the production processes and the inventory decisions, building a direct-to-costumer business prototype [9]. Further ahead, this phenomenal concept of dynamic pricing has been at leads because of the intensified network connections [10]. This has helped both the customers and the sellers under two factors [11], lower menu cost and integrated customer information as a complete database. The internet accessibility has helped buyers or consumers acting as a self-service facility and thus saving time.

The vendors are also benefitted by this concept of dynamic pricing with the amalgamation of web integration and automation in many ways. It eliminates the physical presence of the vendor [12], lowers the input cost, and integrates the customer information under one database and reduces the cost of printing new catalogs [13]. Also it acts not as a one way street between consumers and seller instead an explicit platform to discuss and exchange reviews for better services.

Dynamic pricing as an application can be functional under certain factors, that is, customers' willingness to pay different prices, segmented market availability, less arbitrage potential, fair play rules and revenue cost to be higher than the segmenting and policing capital [14]. Further, it can be implemented in the industries with fixed high cost and low variable cost [10] [15].

Dynamic pricing can be executed through a simple measure of re-pricing the products also according to the competitor's prices. And further reducing the prices during low demand and increasing the prices under high demand scenarios. This process helps in determining the prices accordingly and in improving the seller's profits [16]. Another technical procedure of implementing dynamic pricing is through short term cycles, namely, temporary markdowns and permanent markdowns [7]. Temporary markdown in other words is sale which offers a fixed discount over a fixed period of time, returning back to the original price after sometime. Permanent markdown or clearance is the practice in which the next price of a product can be lower than the current price.

Dynamic pricing has also taken charge in various industries across the globe, as it ameliorates the buying and selling process. The clout of dynamic pricing have been perceived in the industries like, airlines, hotels, electric utilities, retail, internet retail, mobile communication systems, automotive industry, sporting events, car rental companies and insurance sector, to name a few. Another aspect of dynamic pricing which is combinatorial auctions have been implemented in the e-selling, e-procurement, e-logistics, supply chain management and B2B exchange systems [12].

Individual application of dynamic pricing in the airline industry is studied in another name called yield management or revenue management [17]. It involves the process of segmentation of the passengers/travelers under three categories, which are business travelers, casual travelers and hybrid travelers. Another sector using dynamic pricing is automotive industry [9] wherein, the production schedules and inventory decisions are combined together for better profit outcomes and improved supply chain management. The further enhancement provided by dynamic pricing can be stated as easy and quick customer demand presentation along with the upgraded equipment manufacturer status.

To study the determination techniques for dynamic pricing, there are sundry methods available for it. The most familiar and simple method is by survey or observation [18]. The above two methods are implemented by providing a price recommendation function which studies the price to sales ratio and willingness of the consumers to pay, very minutely. Other options like, experimental auctions can also be used to decide optimal pricing strategies. Further, the strategies are also decided according to the kind of market, being mass or niche. Each market needs a skimming and penetration pricing strategy respectively [19].

Five pricing strategies have been discussed to measure dynamic pricing by Magloff [20] namely, Segmented Pricing, Peak Use Pricing, Service Time Pricing, Time of Purchase Pricing, and Changing Conditions Pricing. The segmented pricing involves the price change for the commodity according to the willingness of the customer to pay. Then, the peak user pricing is implemented more often in airline and railway industry where the consumer is charged heavily at peak or rush hours. The service time pricing is a strategy charging high for less service or pre-determined delivery time. Then, time of purchase pricing is executed at the time of purchase when the takeoff time of the flight is less. Lastly, the changing conditions pricing is put forth during a great amount of uncertainty in the market with respect to a product.

Dynamic pricing of the information goods [21] can also be undertaken by following some simple steps, namely,

customization, bundling and versioning. Customization process searches the optimal price for a product and then customizes it accordingly. Bundling includes the selling of two or more products at the same price which helps in reducing cost, billing, administrative cost, consumer valuations and increases the quality thereby by allowing price discrimination to occur. Lastly, versioning as a process discriminates the products and allow consumers to select the preferred one.

Thus, Dynamic Pricing is improving and is becoming an important part of e-commerce industry. Many online and even offline firms are adopting the strategy for Dynamic Pricing in order to attract and retain more customers. This will help in making better strategies for the customers and will eventually help the firm to grow in a better manner. From this argument arise the research question for this study which proposes and evaluates a new form of framework for the various retail organizations to follow. Dynamic pricing based on the combination of mining, statistical and machine learning techniques, is proposed in this paper. Can this combination predict a better pricing for the customers? Better pricing will be defined on the basis of customer's purchase decision.

Rest of the paper is organized as follows. Section 2 discusses existing models in brief which will be helpful in understanding the structure of the dynamic pricing as a concept. Section 3 explains the proposed models which identifies the process through which dynamic pricing is obtained. It will include the explanation of various techniques used in this model. Section 4 presents the data description and results of the proposed models. They are compared with the other techniques used in dynamic pricing for retail stores. Section 5 presents the conclusion, recommendation and future scope for this work.

2. Existing Models

There have been varied models used in today's world to determine dynamic pricing. Some are used to determine the prices for all kinds of products while some are specific in determining a particular cost. The different methods existing and followed in the price determination [12] [17] [22] are as follows.

Agent Based Modeling is a technique that involves factors, agents and rules which analyze the individual or group pricing through various computational methods and actions. *Inventory Based Model* works on the concept of inventory levels and customer service. It can be further sub divided under three categories, Replenishment and Non-replenishment inventory which constitutes the price decision on the basis of the fixed inventory in a given time or the replenishment of the inventory over the time according to the demand and supply. Second sub division is Dependent and Independent demand over time implying the changing customer demands. Third is a Myopic and Strategic customer, stating myopic customers' purchase existing when the price is below their valuation and strategic purchase existing keeping a view of the future price changes.

Data Driven Model involves the price determination through the data collected about the customer preferences and buying patterns. *Game Theory Model* considers more of economics concept and is applied in case there is more number of sellers for a single customer. *Machine Learning Model* incorporates the use of e-market for the understanding of buyers' preferences and patterns and use of algorithms for profit maximization. *Simulation Model* can be used for any decision making model. Also all the other models specified can be used as a simulation model.

Auction Based Model requires six major factors for successful dynamic pricing. First is resource for which auctioning is to be performed, second a defined market structure between the buyers and the sellers, third is preference structure involving product preferences supplied by the agents, fourth is bid structure defining the flexibility of the resource requirement, fifth is market clearing incorporating the matching of supply to demand and last is information feedback signaling price to a bidder to re-define the bid according to the winning bid.

With the various existing models, it was not feasible to combine the effects of more than one model which could solve the problem of Purchase Behavior through Dynamic Pricing in a much more comprehensive way. Thus the current study proposes to develop a framework which would keep the Dynamic Pricing as the base problem to be solved and will determine the appropriate customer segment along with the prediction of most likely purchase range for him. The framework is expected to yield efficient results.

3. Proposed Model

The proposed model considers the amalgamation of three different techniques - to identify the customer segments, appropriate pricing for them, and the prediction for their likely purchase within that price range. The
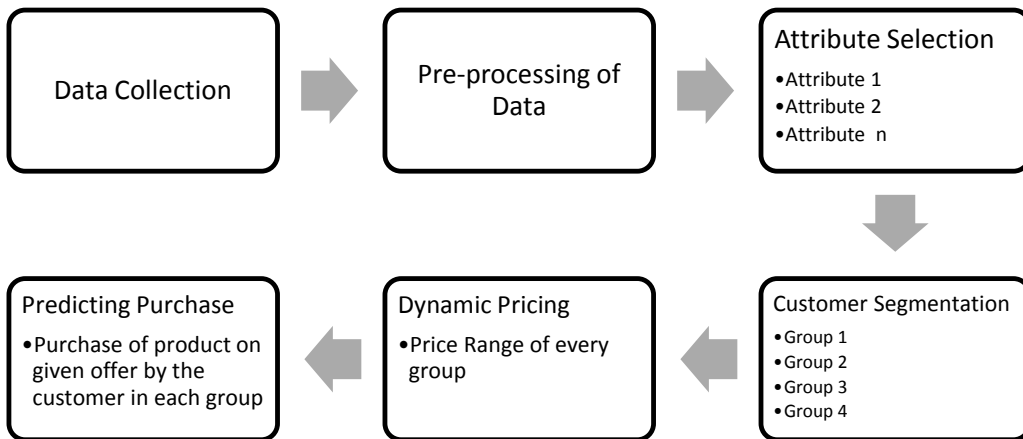
framework is shown in Figure 1.

Figure 1. The proposed framework for predicting online purchase by a customer based on Dynamic Pricing for online

The proposed model consists of following stages for determining the purchase behavior and pricing strategy for the online customers.

(a) Data Collection

This is the first and foremost step in the process of the framework. It involves the collection of data from various data points under an integrated database. For the research purpose, we used a subset of an online marketplace data The schema of the two data sets is shown in Figure 2.

Customer ID	Store Chain	Store Department	Product Category	Product Company	Product Brand	Purchase Date	Product Size	Product Measure	Purchase Quantity	Purchase Amount
Offer ID	Product Category		Product Quantity		Product Company		Offer/price Value		Product Brand	

Figure 2. The first schema corresponds to the transaction database which contained all the customer's transaction records while second schema corresponds to the offer database where different offers in terms of price deductions were mentioned with respect to particular product, category, and company

The transaction database had categorical data for most of the variables as lots of IDs were represented by the variables like customer, chain, department, category, company, brand, size, and measure. Purchase Quantity and amount were found to be continuous variables. Similarly, for offers database, only offer value and product quantity was continuous while others were categorical in nature. A total of 350 million transactions was recorded in the database of around 2.4 million unique customers. For the transactions, we considered only those customers, to whom offers were given and they availed it. This way we had a dataset of all the customers with the different price range while purchasing products from category, company and brand.

(b) Pre-Processing

This step processes all the collected data according to the relevancy they attain for the prediction of price. Preprocessing is also required in order to prepare data sheets for the specific tools for the analysis purpose. The tools used for this study includes R, SAS, and Excel. The data was not present in continuous form and hence new variables were derived to attain a more meaningful data. The derived variables are purchase by offer (POR), purchase by category (PCT), purchase by quantity (PQT), purchase by company (PCY), purchase by brand (PBD), and purchase by channel (PCN). These were calculated by summing up the total purchase amount of a particular customer, counting the total number of offers, category, quantity, company, brand, and channel considered for customer's purchased products, and dividing the two numbers, respectively. The outliers from the data were

removed and data was rolled up for the various analytical tests.

(c) Attribute Selection

This step involves the attributes to be selected, with whose help customer segmentation is carried out. For a new customer, visit attribute, demographic profile, Context, Purchase history, and Purchase intentions should be used as various attributes from the selected data, but in the present research only repeat customers have been considered to find a specific case for this project. For repeat customers, the major variables used are POR, PCT, PQT, PCY, PBD, PCN, purchase amount and purchase quantity. These attributes are used to find similarity amongst the various customers to group the similar type of users.

(d) Customer Grouping

Customer grouping is performed on the basis of selected attributes. K-means clustering algorithm is used to find out similarity amongst the users. The clusters formed for the group are shown in Table 1 and Figure 2. Overall coefficient of variation is found to be 83% which covers up substantial portion of the data set.

Table 1 Clusters formed for the various customers

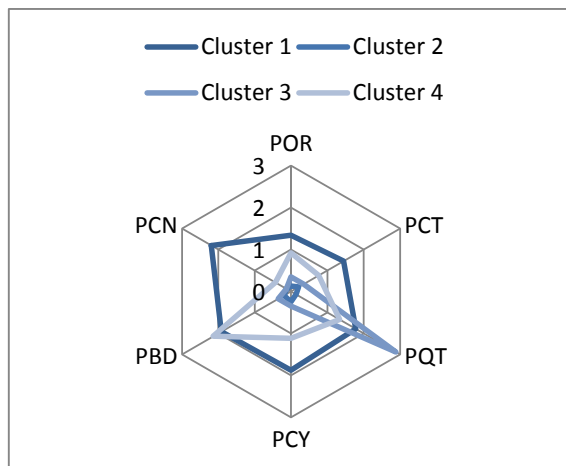| Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Frequency	4688	132307	86255	18129
RMS-SD	1.43	0.48	0.65	0.92
Variable 1 - POR	1.34	0.13	0.35	0.92
Variable 2 - PCT	1.45	0.21	0.35	0.76
Variable 3 - PQT	1.78	0.14	2.89	1.34
Variable 4 - PCY	1.88	0.25	0.35	1.12
Variable 5 - PBD	1.92	0.19	0.35	2.14
Variable 6 - PCN	2.19	0.12	0.13	0.43

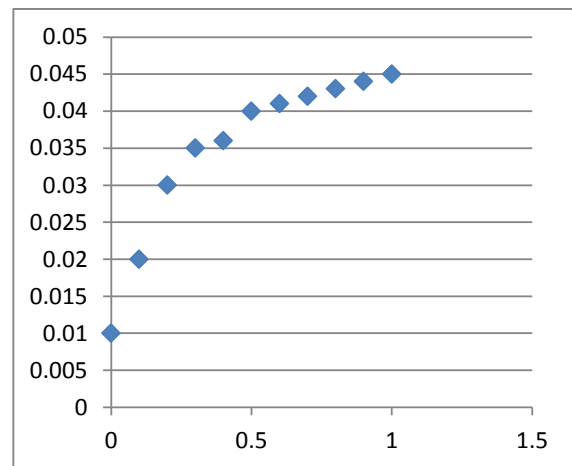
Figure 2 Clusters formed for the various customers

Figure 3 ROC curve for Logistic Regression

(e) Dynamic Pricing

Based on the various segments identified for the customers, the dynamic price range is determined for each of the segment. Dynamic Pricing uses Statistical & Machine Learning techniques to identify the appropriate price range for each segment. Supervised learning is more fruitful as more accuracy can be achieved on the basis of past data. Different price range for different segment will be useful in giving more attention to a particular segment based on its peculiar characteristics. The regression equation formulated for the cluster is as follows:

$$P_i = \beta_0 + \beta_1 POR_i + \beta_2 PCT_i + \beta_3 PQT_i + \beta_4 PCY_i + \beta_5 PBD_i + \beta_6 PCN_i \tag{1}$$

Where, P_i is the price of ith cluster, and β's are the coefficients of the slope and independent variables for every

individual cluster. The price range for any group is depicted through the purchasing power of the customer and thus every cluster will be defined by different price range. Once any customer repeats the purchase from the store, the purchasing power is noticed through historical dataset, and based on his spending and purchasing pattern he is grouped in a specific cluster. Based on the cluster, his price range is predicted and an offer value is depicted to be given to the customer. Table 2 gives the regression results for the four clusters based on purchasing power variables. The overall variation for the clusters is decent enough to consider the price range obtained. And based on these models, the price range of individual customers is obtained.

Table 2 Regression results for price prediction for individual clusters

Variables	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Constant	0.043173*	0.020*	0.0225*	0.0351*
Variable 1 - POR	0.06476*	0.030*	0.03375	0.05265*
Variable 2 - PCT	0.323798*	0.150*	0.16875*	0.26325*
Variable 3 - PQT	0.302211	0.140*	0.1575*	0.2457*
Variable 4 - PCY	0.539663*	0.250	0.28125*	0.43875*
Variable 5 - PBD	0.215865*	0.100*	0.1125	0.1755*
Variable 6 - PCN	2.137064*	0.990*	1.11375	1.73745
R-Square	0.83	0.93	0.91	0.88
Price Range	$300-$500	$1000-$2500	$2000-$5000	$10000-$15000

* Significant at 0.05 level ** Significant at 0.01 level

(f) Predictive Modeling

At this stage, Logistic Regression is used to predict that given an appropriate customer group and an appropriate price range through dynamic pricing methodology, is a customer likely to purchase the product or not. A binary predictor is an appropriate choice given the above framework and helps in identifying the final purchase behavior of the customer. The results for the logistic regression based on purchasing power and price prediction through the multiple regression were calculated on the data set. The complete data was divided into train set and test set in the ratio of 4:1. The area under the curve is depicted in Figure 3.

4. Result Analysis

The framework developed for the price prediction is analyzed for the purchase predictions and the amount of revenue benefits it can produce. For the same product offered at a fixed price for a particular group of customers, our proposed model saw a better revenue generation system with lesser number of errors in predicting customer purchase. The results are shown in Figure 4. The above framework will suitably predict the purchase behavior of the customer. As the time progresses and more data is collected, the supervised learning will produce more accurate results and will be helpful in determining the accurate results for purchase behavior.

5. Conclusion

The proposed framework has been designed using the powerful techniques of Machine Learning, Data Mining and Statistical Methods to predict the purchase behavior of an online customer by selecting an appropriate price range for him based on Dynamic Pricing. This framework has been tested on a large dataset for an e-commerce firm and results are encouraging enough to implement the framework completely. The error rate is reduced and much better price range, which is appropriate for both customer and organization, is being determined. The general framework can be applied in the various industries working in online mode and can be customized to specific applications. The results of the work-in-progress are likely to be discussed in the extension of this study.
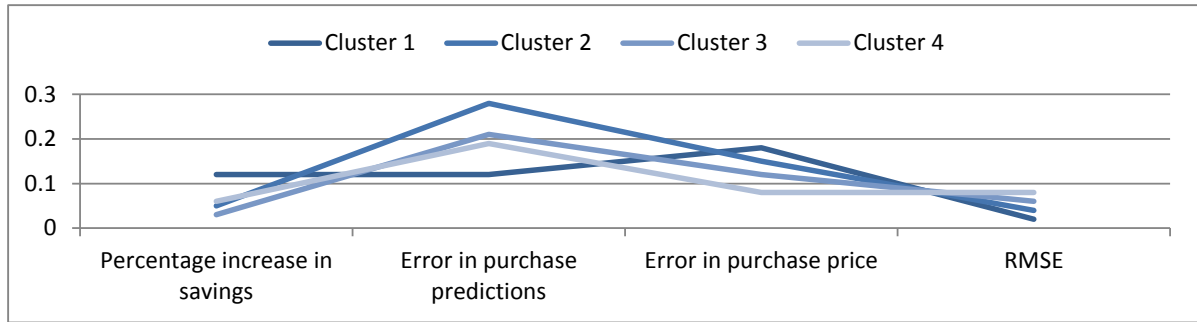
Figure 4 Analysis of the results obtained from the analysis

References

1. Shpanya, A. (2013) "5 Trends To Anticipate In Dynamic Pricing". Retail Touch Points. Accessed on April 1, 2014.
2. Garbarino, E., & Lee, O. F. (2003). Dynamic pricing in internet retail: effects on consumer trust. Psychology & Marketing, 20(6), pp. 495-513.
3. McAfee, R. P., & Te Velde, V. (2006). Dynamic pricing in the airline industry.forthcoming in Handbook on Economics and Information Systems, Ed: TJ Hendershott, Elsevier.
4. Reinartz, W. (2002). Customizing prices in online markets. Management (www. unimib. it/symphonya), (1), pp. 55-65.
5. Strauss, J., Frost, R., & Ansary, A. I. (2009). E-marketing. Pearson Prentice Hall.
6. Cummings, T. (2013). Csmonitor.com, Everything you need to know about dynamic pricing, accessed from http://www.csmonitor.com/Business/Saving-Money/2013/1104/Everything-you-need-to-know-about-dynamic-pricing on 19th June, 2014.
7. Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. Management Science, 49(10), pp. 1287-1309.
8. Fitkov-Norris, E. D., & Khanifar, A. H. M. A. D. (2000). Dynamic pricing in mobile communication systems.
9. Biller, S., Chan, L. M. A., Simchi-Levi, D., & Swann, J. (2005). Dynamic pricing and the direct-to-customer model in the automotive industry. Electronic Commerce Research, 5(2), pp. 309-334.
10. Weiss, R. M., & Mehrotra, A. K. (2001). Online dynamic pricing: efficiency, equity and the future of e-commerce. Va. JL & Tech., 6, pp. 11-17.
11. Dolan, R. J., & Moon, Y. (2000). Pricing and market making on the Internet.Journal of Interactive Marketing, 14(2), pp. 56-73.
12. Narahari, Y., & Dayama, P. (2005). Combinatorial auctions for electronic business. In Sadhana Academy Proceedings in Engineering Sciences, 30(2-3), pp. 179-211. Indian Academy of Sciences.
13. Rutherford, D. (1995). Routledge dictionary of economics. Taylor & Francis.
14. Randel, S. (2009). Dynamic Pricing: An old practice is making a comeback online, iProspect.com accessed from http://www.iprospect.com/blog/featured/dynamic-pricing-an-old-practice-is-making-a-comeback-online.html on 12th May, 2014.
15. Varian, H. R., & Shapiro, C. (1999). Information rules: a strategic guide to the network economy. Harvard Business School Press, Cambridge.
16. Dasgupta, P. R., & Hashimoto, Y. (2004). Multi-attribute dynamic pricing for online markets using intelligent agents. In Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, 1, pp. 277-284. IEEE Computer Society.
17. Narahari, Y., Raju, C. V. L., Ravikumar, K., & Shah, S. (2005). Dynamic pricing models for electronic business. In Sadhana (Academy Proceedings in Engineering Sciences, 30(2-3), pp. 231-256). Indian Academy of Sciences.
18. Lippert, J. (2011). Dynamic Price Optimisation for the retail market, Prudysys AG.
19. Bergemann, D., & Välimäki, J. (2006). Dynamic pricing of new experience goods. Journal of Political Economy, 114(4), pp. 713-743.
20. Magloff, L. (2014). Dynamic Pricing Strategy, Chron.com, Small business accessed from http://smallbusiness.chron.com/dynamic-pricing-strategy-5117.html on 12th May, 2014.
21. Viswanathan, S., & Anandalingam, G. (2005). Pricing strategies for information goods. Sadhana, 30(2-3), pp. 257-274.
22. Boucek, C. H., & Conway, T. E. (2007). Dynamic pricing analysis. JSTOR. WPI, 1.
23. Kaggle Database, Acquire value shoppers challenge, Online retail customers. Accessed from https://www.kaggle.com/c/acquire-valued-shoppers-challenge/data on 15th april 2014.