Customer lifetime value model in an online toy store

B. Nikkhahan \(^1\); A. Habibi Badrabadi \(^2\); M.J. Tarokh \(^3*\)

\(^1\) Postgraduate student, IT Group, Dep. of Industrial Engineering, K.N. Toosi University of Technology, Tehran, Iran

\(^2\) Associate Professor, IT Group, Dep. of Industrial Engineering, K.N. Toosi University of Technology, Tehran, Iran

Abstract: Business all around the world uses different approaches to know their customers, segment them and formulate suitable strategies for them. One of these approaches is calculating the value of each customer for the company. In this paper by calculating Customer Lifetime Value (CLV) for individual customers of an online toy store named Alakdolak, three customer segments are extracted. The level of profitability for customers is identified, and finally suitable marketing strategies for each segment are developed. The results indicate that the company should increase its low price products and develop special programs for those that buy high price products and have high loyalty. Logistic regression as a data mining technique is used to present the customer defection and future purchase probability models and for each model, verification and validation is done.

Keywords: Customer lifetime value; Data mining; Online store

1. Introduction

In the past two decades, the firms tended to focus on either cost management or revenue growth. When a firm adopts one of these approaches it looses out on the other (Rust \textit{et al.}, 2004). Assessing the value of individual customers and employing customer level strategies based on customers’ worth to the firm, leads to effective allocation of resources and efforts across profitable customers. Many customer oriented firms realize that the customers are valued more than the profit they bring in every transaction. Customers’ value has to be based on their contribution to the firm across the duration of their relationship with the firm. In simple terms, the value of a customer is the value the customer brings to the firm over his/her lifetime (Kumar 2006).

Valuing customers is a central issue of any commercial activity. The value of an individual customer is important for the detection of the most valuable ones, which deserve to be closely followed, and for the detection of the less valuable ones, to which the company should pay less attention. At the aggregated level, a marketing campaign targeting a group of customers can be budgeted more efficiently when the value of this group is known. Customers are an important asset, and as such, have to be precisely valued.

Customer Lifetime Value (CLV) plays a major role in customer valuation, in particular Churn Analysis and Retention Campaign Management. In the context of Churn Analysis, the CLV of a customer or a segment is important complemen-
towards purchasing at the online store, revealing significant, direct influences from usefulness, enjoyment, trustworthiness and settlement performance. Da Silva and Syed Alwi (2008) found that factors such as ease of use, “personalization”, security and customer care are significant in determining the corporate brand image of the online e-tailer.

2. Fundamentals of CLV modeling

There are some definitions for CLV in Table 1. CLV is generally defined as the present value of all future profits obtained from a customer over his or her life of relationship with a firm. A simple mathematical model of CLV for a customer (omitting customer subscript) is (Tarokh et al., 2006; Gupta et al., 2004; Reinartz and Kumar, 2003; Sheth et al., 2002):

\[ CLV = \sum_{i=1}^{n} \frac{R_i - C_i}{(1 + d)^i} \]

where \( i \) is period of cash flow from customer transactions; \( R_i \) the revenue from customer in period \( i \); \( C_i \) the total cost of generating the revenue \( R_i \) in period \( i \); \( d \) interest rate; and \( n \) the total number of periods of projected life of customer under consideration. Therefore the numerator is the net profit that has been obtained at each period while the denominator transforms the profit into the current value. This model is the most basic model that ignores the function of sales and costs. Expanding the model with data mining techniques can consider future customers’ purchase behavior.

Researchers either build separate models for customer acquisition, retention, and margin or sometimes combine two of these components. For example, Thomas (2001) and Reinartz et al. (2005) simultaneously captured customer acquisition and retention. Fader et al. (2005) captured recency and frequency in one model and built a separate model for monetary value. However, the approaches for modeling these components or CLV differ across researchers. Gupta et al. (2006) described six modeling approaches:

- RFM models,
- Probability models,
- Econometric models,
- Persistence models,
- Computer science models,
- Diffusion / Growth models.

RFM models create “cells” or groups of customers based on three variables: Recency, Frequency, and Monetary value of their prior purchases. Several recent studies have compared CLV models (discussed later) with RFM models and found CLV models to be superior. Reinartz and Kumar (2003) used a catalog retailer’s data of almost 12,000 customers over 3 years to compare CLV and RFM models. They found that the revenue from the top 30% of customers based on the CLV model was 33% higher than the top 30% selected based on the RFM model.

Venkatesan and Kumar (2004) also compared several competing models for Customer Selection. Using data on almost 2,000 customers from a business-to-business (B2B) manufacturer, they found that the profit generated from the top 5% customers as selected by the CLV model was 10% to 50% higher than the profit generated from the top 5% customers from other models (e.g., RFM, past value, etc.).

A probability model is a representation of the world in which observed behavior is viewed as the realization of an underlying stochastic process governed by latent (unobserved) behavioral characteristics, which in turn vary across individuals. Schmittlein and Peterson (1994), Colombo and Jiang (1999), Reinartz and Kumar (2000, 2003), Fader et al. (2004) have all proposed probability models to compute CLV.

Table 1: LTV definitions.

<table>
<thead>
<tr>
<th>LTV definitions</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>The NPV of all future contributions to overhead and profit.</td>
<td>Roberts and Berger (1989)</td>
</tr>
<tr>
<td>NPV of a future stream of contributions to overheads and profits.</td>
<td>Jackson (1994)</td>
</tr>
<tr>
<td>The NPV of all future contributions to profit and overhead expected from the customer.</td>
<td>Courtheoux (1995)</td>
</tr>
<tr>
<td>The NPV of the stream of contributions to profit that results from customer transactions and contacts.</td>
<td>Pearson (1996)</td>
</tr>
<tr>
<td>Expected profit from customers, exclusive of cost related to customer management.</td>
<td>Blattberg and deighton (1996)</td>
</tr>
<tr>
<td>The total discounted net profit that a customer generates during her life on the house list.</td>
<td>Bitran and Mondshein (1996)</td>
</tr>
<tr>
<td>The present value of all future profits generated from a customer.</td>
<td>Gupta and Lehmann (2003)</td>
</tr>
<tr>
<td>Sum of cumulated cash flows - discounted using the Weighted Average Cost of Capital (WACC) - of a customer over his or her entire lifetime with the company.</td>
<td>Kumar (2006)</td>
</tr>
<tr>
<td>Estimated monetary value that the client will bring to the firm during the entire lifespan of his/her commercial relationship with the company, discounted to today's value.</td>
<td>Yamamoto Sublaban and Aranha (2008)</td>
</tr>
</tbody>
</table>
Econometric models deal with customer acquisition, retention, and expansion (cross-selling or margin) and then combine them to estimate CLV. Like econometric models of CLV, persistence models focus on modeling the behavior of its components, that is, acquisition, retention, and cross-selling. When sufficiently long-time series are available, it is possible to treat these components as part of a dynamic system. To date, this approach, known as persistence modeling, has been used in a CLV context to study the impact of advertising, discounting, and product quality on customer equity (Yoo and Hanssens 2005) and to examine differences in CLV resulting from different customer acquisition methods (Villanueva et al., 2006).

The vast computer science literature in data mining, machine learning, and nonparametric statistics has generated many approaches that emphasize predictive ability. These include projection-pursuit models; neural network models; decision tree models; spline-based models such as Generalized Additive Models (GAM), Multivariate Adaptive Regression Splines (MARS), Classification and Regression Trees (CART); and Support Vector Machines (SVM).

CLV is the long-run profitability of an individual customer. This is useful for Customer Selection, Campaign Management, Customer Segmentation, and Customer Targeting (Kumar 2006b). Whereas these are critical from an operational perspective, CLV should be aggregated to arrive at a strategic metric that can be useful for senior managers. With this in mind, several researchers have suggested that we focus on Customer Equity (CE), which is defined as the CLV of current and future customers (Blattberg et al., 2001; Gupta and Lehmann 2005; Rust et al., 2004).

3. Improved CLV model

In spite of simple formulation (Equation 1), researchers have used different variations in modeling and estimating CLV. Some researchers have used an arbitrary time horizon or expected customer lifetime (Reinartz and Kumar 2000; Thomas 2001), whereas others have used an infinite time horizon (e.g., Fader et al., 2005; Gupta et al., 2004). Gupta and Lehmann (2005) showed that using an expected Customer Lifetime generally overestimates CLV, sometimes quite substantially. Calculating LTV with the basic model has some deficiencies:

- It cannot consider customer churn probability,
- It cannot consider factors as cross-selling opportunity,
- The possibility of up-selling cannot be calculated.

Hwang (2003) suggested a complimentary model (Equation 2):

\[
LTV_i = \sum_{t=1}^{N_i} \pi_p(t_i)(1+d)^{N_i-t_i} + \sum_{t=N_i+1}^{N} \frac{\pi_f(t_i) + B(t_i)}{(1+d)^{t_i-N_i}}
\]

where:

- \(T_i\) Service period index of customer \(i\).
- \(N_i\) Total service period of customer \(i\).
- \(d\) Interest rate.
- \(E(i)\) Expected service period of customer.
- \(\pi_p(t_i)\) Past profit contribution of customer \(i\) at period \(t_i\).
- \(\pi_f(t_i)\) Future profit contribution of customer \(i\) at period \(t_i\).
- \(B(t_i)\) Potential benefits from customer \(i\) at period \(t_i\).
- \(\pi_p(t_i)\) Can be calculated from Equation (1).
- \(\pi_f(t_i)\) Is the prediction of \(\pi(t_i)\) in the future and can be calculated using regression techniques.

In order to calculate \(E(i)\) we should describe the relation between expected service period of customers and customer defection (churn) rate. Let \(y\) be the number of service time required until a customer service is stopped and let \(P_{Churn}\) be the churn probability of a customer. Then, the probability of mass function of service period is given by:

\[
P(n) = P(y = n) = P_{Churn} \times (1-P_{Churn})^{n-1}, n = 1, 2, 3, \ldots
\]

\[
\therefore y \sim \text{Geo}(P_{Churn}) \Rightarrow E[i] = \frac{1}{P_{Churn}}
\]
In order to calculate \( B(t_i) \), we should use Equation (4) as follows:

\[
B(t_i) = \sum_{j=1}^{k} \text{prob}_{ij} \times \text{prof}_{ij} \quad k = 1, 2, 3,
\]

= Potential benefits of customer in period \( i \) \hspace{1cm} (4)

where \( \text{prob}_{ij} \) is the probability that customer \( i \) would use the service \( j \) among \( k \)-potential services. \( \text{prof}_{ij} \) means the profit that a company can get from customers \( i \) who uses optional service \( j \). It is available by substituting the cost of each optional service from the charge of each optional service. However, we should use data mining techniques in order to calculate \( \text{prob}_{ij} \).

4. Case study: Online toy store

In this study, data were gathered from an online store. Alakdolak is an online store that sells toys. Children are its customers but most of the time other people buy from the store for their relatives. Children are indirect customers, but we need their age to predict their relatives’ behavior in online store. So we consider children as customers. Our data includes two types of data. One of them is personal information like age, sex, relation of the buyer to the customer (like son, daughter). Another one is the shopping history of the customers like the name of the product, its cost and the date of shopping.

The data are related to 182 customers and their 1935 purchasing records in the last five months of 2008. First we need to calculate the probability of customer churn and the probability of purchasing each product type by the customer. Logistic regression was used as a data mining technique. A table of 182 customers was used with 11 Data fields. The churn behavior of customers recorded as a binary variable. Equation (5) represents the logistic regression (Menard and Scot 1995). Where \( \beta_0 \) is the constant value and \( \beta_i \) is model parameters.

\[
P(y = 1) = \frac{1}{1 + \exp^{-\left(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \right)}} \hspace{1cm} (5)
\]

5. Modeling and data mining

So they must have a definition for a defected customer. For this purpose they set up an interview with the managers of the store. They believed that customers of this online store are children and they have limited period of using toys and most of the parents buy toys for their children in their childhood. So if they didn't buy any products from the shop for three months, we can assume that they are buying from other stores.

We can categorize the products into three categories: Low, Medium and High Price. Now we want to calculate the probability of buying these products by customers. Based on the case the researchers present four models: A model for calculating the customer defection probability, and 3 models for calculating cross-selling probability.

5.1. Calculating customer defection probability

The researchers used to SPSS software to establish logistic regression model. Customer churn data were chosen as the response variable and backward stepwise method with Wald statistics was implemented. In the first step, all variables entered into the model and then through hypothesis test \( H_0: \beta_i = 0 \) some of them eliminated. In the first model, \( B \) is independent variable coefficient. Using Microsoft Excel, the probability of customer churn was calculated for each customer.

5.2. Calculating cross-selling probability

All the 3 types of products were considered for cross-selling opportunity. It consists of 3 models (one model for each product type) with SPSS software as follows:
Model 1: Prediction of customer churn.

Variables in the Equation

<table>
<thead>
<tr>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>df</th>
<th>Sig</th>
<th>Exp(B)</th>
</tr>
</thead>
</table>

Step 1
- Sex: 0.699, p = 0.197, Exp(B) = 1.912
- Related: -0.523, p = 0.589, Exp(B) = 0.59
- Age: 0.017, p = 0.801, Exp(B) = 1.017
- Sale count: -2.381, p = 0.020, Exp(B) = 0.10
- Day count: 1.868, p = 0.033, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Step 2
- Sex: 0.712, p = 0.187, Exp(B) = 2.039
- Related: 1.317, p = 0.037, Exp(B) = 3.734
- Day count: 1.868, p = 0.022, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Step 3
- Day count: 1.868, p = 0.022, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Model 2: Prediction of buying the low price products.

Variables in the Equation

<table>
<thead>
<tr>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>df</th>
<th>Sig</th>
<th>Exp(B)</th>
</tr>
</thead>
</table>

Step 1
- Sex: 0.731, p = 0.173, Exp(B) = 2.077
- Related: 1.317, p = 0.037, Exp(B) = 3.734
- Day count: 1.868, p = 0.022, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Step 2
- Day count: 1.868, p = 0.022, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Model 3: Prediction of buying the medium price products.

Variables in the Equation

<table>
<thead>
<tr>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>df</th>
<th>Sig</th>
<th>Exp(B)</th>
</tr>
</thead>
</table>

Step 1
- Sex: 0.456, p = 0.499, Exp(B) = 1.56
- Related: -0.523, p = 0.589, Exp(B) = 0.59
- Age: 0.017, p = 0.801, Exp(B) = 1.017
- Sale count: -2.381, p = 0.020, Exp(B) = 0.10
- Day count: 1.868, p = 0.033, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Step 2
- Day count: 1.868, p = 0.022, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Model 4

Variables in the Equation

<table>
<thead>
<tr>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>df</th>
<th>Sig</th>
<th>Exp(B)</th>
</tr>
</thead>
</table>

Step 1
- Sex: 0.456, p = 0.499, Exp(B) = 1.56
- Related: -0.523, p = 0.589, Exp(B) = 0.59
- Age: 0.017, p = 0.801, Exp(B) = 1.017
- Sale count: -2.381, p = 0.020, Exp(B) = 0.10
- Day count: 1.868, p = 0.033, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Step 2
- Day count: 1.868, p = 0.022, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Model 5

Variables in the Equation

<table>
<thead>
<tr>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>df</th>
<th>Sig</th>
<th>Exp(B)</th>
</tr>
</thead>
</table>

Step 1
- Sex: 0.456, p = 0.499, Exp(B) = 1.56
- Related: -0.523, p = 0.589, Exp(B) = 0.59
- Age: 0.017, p = 0.801, Exp(B) = 1.017
- Sale count: -2.381, p = 0.020, Exp(B) = 0.10
- Day count: 1.868, p = 0.033, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Step 2
- Day count: 1.868, p = 0.022, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Model 6

Variables in the Equation

<table>
<thead>
<tr>
<th>B</th>
<th>SE</th>
<th>Wald</th>
<th>df</th>
<th>Sig</th>
<th>Exp(B)</th>
</tr>
</thead>
</table>

Step 1
- Sex: 0.456, p = 0.499, Exp(B) = 1.56
- Related: -0.523, p = 0.589, Exp(B) = 0.59
- Age: 0.017, p = 0.801, Exp(B) = 1.017
- Sale count: -2.381, p = 0.020, Exp(B) = 0.10
- Day count: 1.868, p = 0.033, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32

Step 2
- Day count: 1.868, p = 0.022, Exp(B) = 6.475
- Sale product cost sum: -1.000, p = 0.319, Exp(B) = 0.32
6. Model verification

There are several statistics which can be used for comparing alternative models or evaluating the performance of a single model (Menard and Scott, 1995).

6.1. Model Chi-square

In logistic regression, the statistic is analogous to the F-test in linear regression (Menard and Scott, 1995). By multiplying the log-likelihood by -2 it approximates a distribution (Menard and Scott, 1995).

\[ G_M = D_0 - D_M \]  

where \( D_0 \) is initial log likelihood and \( D_M \) is final log likelihood. \( D_0 \) is analogous to the total sum of square (SST) in linear regression analysis and \( D_M \) is analogous to the Error sum of square (SSE).

\[ G_M \sim X^2 \]  

where \( k \) is degree of freedom and is equal to the number of remaining variables in the final step of modeling. For example for model 1, \( G_M = 225.744 \) and because \( G_M > X^2_{0.99, 2} = 9.210 \) the null hypothesis is rejected and model significance is verified.

7. Model accuracy

SPSS creates classification table for each step in the logistic regression model. Each classification table summarizes the observed and predicted values to interpolate predictive efficiency/ accuracy for each step in the model. The bigger the percentage correct the better the model. The classification table is presented following each Model 1-4.

7.1. Pseudo- \( R^2 \) statistics

It is the proportion of the variance in the dependent variable which is explained by the variance in the independent variables. There are several Pseudo- \( R^2 \) statistics. Analogous to \( R^2 = \frac{SSR}{SST} \) for linear regression is Pseudo- \( R^2 \) in logistic regression (Menard and Scott, 1995). Pseudo- \( R^2 \) indicates how much the inclusion of the independent variable in the model reduces the badness of the fit. It varies between 0 (independent variable is useless in the prediction of the dependent variable) and 1 (Independent variable in the model predicts the dependent variable perfectly) and is calculated by dividing the model chi-square (\( G_M \)) by the initial log-likelihood function (Menard and Scott, 1995).

\[ \text{Pseudo -} R^2 = \frac{G_M}{G_M + D_M} \]
Pseudo-$R^2$ is calculated for each Model 1-4 and variables eliminated by SPSS at each step are tested. For example conclusion for Model 4 (High Price) is shown in Table 5.

Table 2: R$^2$ changes for churn model (Model 1).

<table>
<thead>
<tr>
<th>Step</th>
<th>-2 Log Likelihood</th>
<th>Cox &amp; Snell R$^2$</th>
<th>Nagelkerke R$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24.109</td>
<td>.715</td>
<td>.953</td>
</tr>
<tr>
<td>2</td>
<td>24.281</td>
<td>.714</td>
<td>.952</td>
</tr>
<tr>
<td>3</td>
<td>24.730</td>
<td>.714</td>
<td>.951</td>
</tr>
<tr>
<td>4</td>
<td>25.563</td>
<td>.712</td>
<td>.950</td>
</tr>
<tr>
<td>5</td>
<td>26.562</td>
<td>.711</td>
<td>.948</td>
</tr>
</tbody>
</table>

a. Estimation terminated at iteration number 11 because parameter estimates changed by less than .001.

b. Estimation terminated at iteration number 10 because parameter estimates changed by less than .001.

c. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

Table 3: R$^2$ changes for low price model (Model 2).

<table>
<thead>
<tr>
<th>Step</th>
<th>-2 Log Likelihood</th>
<th>Cox &amp; Snell R$^2$</th>
<th>Nagelkerke R$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>91.380</td>
<td>.587</td>
<td>.783</td>
</tr>
<tr>
<td>2</td>
<td>91.493</td>
<td>.587</td>
<td>.782</td>
</tr>
<tr>
<td>3</td>
<td>92.632</td>
<td>.584</td>
<td>.779</td>
</tr>
<tr>
<td>4</td>
<td>94.507</td>
<td>.580</td>
<td>.773</td>
</tr>
<tr>
<td>5</td>
<td>97.723</td>
<td>.572</td>
<td>.763</td>
</tr>
</tbody>
</table>

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

c. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

d. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

e. Estimation terminated at iteration number 2 because parameter estimates changed by less than .001.

Table 4: R$^2$ changes for medium price model (Model 3).

<table>
<thead>
<tr>
<th>Step</th>
<th>-2 Log Likelihood</th>
<th>Cox &amp; Snell R$^2$</th>
<th>Nagelkerke R$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>235.959</td>
<td>.086</td>
<td>.115</td>
</tr>
<tr>
<td>2</td>
<td>236.053</td>
<td>.085</td>
<td>.114</td>
</tr>
<tr>
<td>3</td>
<td>237.013</td>
<td>.051</td>
<td>.107</td>
</tr>
<tr>
<td>4</td>
<td>238.890</td>
<td>.071</td>
<td>.095</td>
</tr>
<tr>
<td>5</td>
<td>241.419</td>
<td>.058</td>
<td>.077</td>
</tr>
<tr>
<td>6</td>
<td>242.526</td>
<td>.052</td>
<td>.070</td>
</tr>
</tbody>
</table>

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

b. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

c. Estimation terminated at iteration number 2 because parameter estimates changed by less than .001.

Table 5: R$^2$ changes for high price model (Model 4).

<table>
<thead>
<tr>
<th>Step</th>
<th>-2 Log Likelihood</th>
<th>Cox &amp; Snell R$^2$</th>
<th>Nagelkerke R$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>39.393</td>
<td>.690</td>
<td>.919</td>
</tr>
<tr>
<td>2</td>
<td>39.525</td>
<td>.689</td>
<td>.919</td>
</tr>
<tr>
<td>3</td>
<td>39.719</td>
<td>.689</td>
<td>.919</td>
</tr>
<tr>
<td>4</td>
<td>40.773</td>
<td>.687</td>
<td>.916</td>
</tr>
<tr>
<td>5</td>
<td>41.514</td>
<td>.686</td>
<td>.915</td>
</tr>
</tbody>
</table>

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.
8. Model validation

Lift chart is one of performance test techniques (Hwang, 2003). The lift value is the ratio of the number of customers who actually left the company within the certain section to the total number of customers who left when we divided sections by sorting target customers in ascending order of high churn rate. The lift chart as proposed above generally shows the cumulative percentage of deviated customers. The researchers conclude that the reliability of the model is good when the graph is skewed to the left. In some notes the base line is mentioned as the measure of reliability (Hwang, 2003). As showed in Figures 1 to 4, "churn model", "low price model", and "high price model" have better results unlike "medium price model". So we can conclude that the model of buying medium price products does not reflect the actual behavior of customers truly.

9. Discussion

Through logistic regression modeling, we can calculate Equation (2) and consider three types of value for customers. As customer life cycle has been considered for five months and we have traced customer purchase behavior in five months, the present value of customers concluded without discount rate. The present value of customers for the online store is sum of their shopping cost. Customer future value is calculated with Equation (4), considering logistic regression Models 2, 3, and 4. We consider customer loyalty as another measure to segment the customers. We have calculated the customer churn probability through Model 1. So we can define:

\[ \text{Customer Loyalty} = 1 - \text{Churn probability} \]

To distinguish between customer segments we used a 3D diagram showed in Table 4.

As shown in Figures 5 and 6, we can categorize the customers of the online store in three main segments:

1. Those that their current value, future value and loyalty are below the average,
2. Those that their current value and future value are below the average, and their loyalty is above the average,
3. Those that their current value, future value and loyalty are above the average.

First customer segment consists of those that their frequency and cost of buying are low. Most of them shopped from the site just one time and they never came back.

Customers in the second segment, generate most of the customer value of the store. Most of them bought cheap products, but their loyalty is high.

Figure 5: Customer segments.
Customer lifetime value model in an online toy store

Figure 6: Future expected value/customer churn rate.

Table 6: Segmentation of customers based on Customer Lifetime Profits and relationship duration (Reinartz et al., 2002).

<table>
<thead>
<tr>
<th>BUTTERFLIES</th>
<th>TRUE FRIENDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Customer Lifetime Profits</td>
<td></td>
</tr>
<tr>
<td>• Good fit between company’s offerings and customers’ needs</td>
<td>• Excellent fit between company’s offerings and customers’ needs</td>
</tr>
<tr>
<td>• High Profit potential</td>
<td>• Highest profit potential</td>
</tr>
<tr>
<td>• Action</td>
<td>• Action</td>
</tr>
<tr>
<td>o Aim for transactional satisfaction, not attitudinal loyalty</td>
<td>o Consistent intermittent spaced communication</td>
</tr>
<tr>
<td>o Maximize profits from these accounts as long as they are active</td>
<td>o Achieve attitudinal and behavioral loyalty</td>
</tr>
<tr>
<td>o Stop investing once inflection point is reached</td>
<td>o Invest to nurture/defend/retain</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STRANGERS</th>
<th>BARNACLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Customer Lifetime Profits</td>
<td></td>
</tr>
<tr>
<td>• Little fit between company’s offerings and customers’ needs</td>
<td>• Limited fit between company’s offerings and customers’ needs</td>
</tr>
<tr>
<td>• Lowest profit potential</td>
<td>• Low profit potential</td>
</tr>
<tr>
<td>• Action</td>
<td>• Action</td>
</tr>
<tr>
<td>o Make no investment in these relationships</td>
<td>o Measure size and share of wallet</td>
</tr>
<tr>
<td>o Make profit on every transaction</td>
<td>o If share-of-wallet is low, focus on specific up and cross selling</td>
</tr>
<tr>
<td></td>
<td>o If size of wallet is small, impose strict cost controls</td>
</tr>
</tbody>
</table>

| Low Relationship Duration | High |
We expect that they will buy products from the store again in the future. This customer segment approximately consists of half of the store's customers. Third customer segment are store's special customers. This segment and the first segment are approximately equal in amount.

They are generating more value than the others for the store; we can expect them to generate this value in the future, because they have high future value. In spite of this, because of their loyalty, the store can sell more products to this customer segment. Reinartz et al. (2002) presented a segmentation of customers based on Customer Lifetime Profits and relationship duration. This model is used to develop suitable marketing strategies for the store. As shown in Table 6 customers are segmented by their value for the firm and their relationship duration. We can locate our segments on this table. We categorized the customers by their loyalty and their value for the store. So our first segment includes strangers, second segment includes barnacles and third segment includes true friends, and we don't have any butterflies.

The company must increase low price products because most of the store's profits belong to the barnacles. Each one of them generates small value, but they have high loyalty and they are more than 50 percent of the customers. So totally, it is
the most valuable customer segment for the company. The true friends are special customers of the company. They buy expensive products and they have high loyalty, but they are a small portion of the customers. As showed in Table 6, the company should develop special programs to have better interaction with them. Most of the strangers buy just one time and they generate small values. The company must increase the number of true friends and barnacles; as a result it leads to decrease the number of strangers.

An overview of all tasks that are performed in this study is presented in Figure 7. At the first step, raw data must become ready for building the models. In the preprocessing of data, many tasks are done. Some concepts are defined such as customer churn, low, medium and high price products. For example a defected customer is one that didn’t buy anything from the store since three month ago. Another task in preprocessing is converting raw data to useful date. For example deriving number of purchases of each customer, sum of its purchases, number of days from the last purchase and so on. After making the data ready, they imported to SPSS software and with linear regression, four models were built. Then we verified and validated the models. In validation process we found out that medium price model is not valid, but the others are valid.

In order to calculate the future value of each customer, we need his or her present value and probability of defection. So for each customer, future value was calculated. CLV of customer calculated by their present and future values. Then the customers segmented by their loyalty, present and future values. Finally based on the segmentation suitable strategies are generated to dealing with different customer segments.

10. Conclusion

Businesses want to allocate their scarce resources to their customers effectively. They want to evaluate and categorize different segments of customers, and then formulate suitable strategies for them. In this paper, Customer Lifetime Value (CLV) is used for this purpose. At first, CLV and its different calculating approaches are described. Then it used to analyze the customers of an online store called “Alakdolak”. To calculate CLV, the probability of customer defection and the probability of buying three kinds of product named "low price", "medium price" and "high price" are calculated. Linear regression is used as a data mining technique for these models and verification and validation is done for each model. The validation process revealed that unlike the other models, "medium price model" does not reflect the actual behavior of customers truly.

Finally with calculating CLV, customers of the store are categorized into three different segments: strangers, barnacles, and true friends. The results showed that barnacles create most of the values of the company.

So company can increase the number of low price products to increase its profit. True friends spend more money in the store and have high loyalty. So the company should extend its interactions with these customers to increase the number of them, and decrease the number of strangers that buy just one time.

References


Friedman, J. H, (2003), *Recent advances in predictive (machine) learning*. Working paper, Department of Statistics, Stanford University, Stanford, CA.


Reinartz, W.; Thomas, J. and Kumar, V., (2005), Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing*, 69(1), 63-79.


Yoo, S. and Hanssens, D. M., (2005), Modeling the sales and customer equity effects of the marketing mix. Working paper, University of California, Los Angeles, Anderson School of Management.