

Predictive Analytics for Increased Loyalty and Customer Retention in Telecommunication Industry

Oladapo K. A.

Department of Computer
Science, School of Computing
and Engineering (CES)
Babcock University, Ilesha-
Remo, Ogun State, Nigeria

Omotosho O. J.

Department of Computer
Science, School of Computing
and Engineering (CES)
Babcock University, Ilesha-
Remo, Ogun State, Nigeria

Adeduro O. A.

Department of Computer
Science, School of Computing
and Engineering (CES)
Babcock University, Ilesha-
Remo, Ogun State, Nigeria

ABSTRACT

Literature has indicated that to engage a new customer cost at least 6 – 10 times higher than retaining the existing ones. The competitive nature of the telecommunication industry has made customer retention to be a crucial responsibility for telephone services provider. Since customer retention is a vital element for every establishment to be conscious of in retaining loyal customers, so also is the ability to perfectly predict customer retention is very necessary. Customer retention prediction models are highly needed by the telecommunication industry to efficiently manage the retention of existing customers. This paper proposes a logistic regression model to predict customer retention in the telecommunication industry. The results indicate that logistic regression can predict customer retention with the accuracy of 95.5%. Furthermore, it was observed that when billing issues are resolved it is more likely to retain customer while value-added service and short message service issues are associated with the likelihood of exhibiting customer retention.

General Terms

Predictive Modelling, Predictive Analytics, Data Mining Techniques, Information System.

Keywords

Prediction, Predictive Analytics, Loyalty, Customer Retention

1. INTRODUCTION

The telecommunication industry has more competition than every other industry which gives the customers more choices. To acquire a new customer costs at least 6 – 10 times more than retaining the existing ones (Pogol, 2007). To every industry the cost acquiring new customers takes the large sum of the administrative expenses; consequently, customer loyalty and retention turns into the major factors to be considered in maintaining growth knowing fully well that industries in a competitive market like telecommunication majorly depends on the ever-growing profits that come from the existing loyal customers. Therefore efforts are to be made to increase loyalty and retain customers. It is of high benefit to the industry to build a more effective customer retention prediction model.

This paper suggests logistic regression model design, a good model from acquired customer data to predict retention. The remaining part of the paper is organized as follows. Section 2 reviews literature in relation to customer loyalty and retention.

Section 3 proposes a model for customer retention and section 4 presents the results of the acquired data set, finally, section 5 provides the conclusion.

2. LITERATURE REVIEW

2.1 Customer Loyalty

Loyalty is seen as a concept which has its base from the customer behaviour theory and reflects what customer may portray to services, activities or brand (Boohene and Agyapong, 2011). The concept has been identified as part of main sources of industry scores and competitive advantage. Industry growth and profit are primarily stimulated by customer's loyalty (Heskett, Jones, Loveman, Sasser, & Schlesinger, 2008). Customer loyalty is the willingness of a customer to retain his/her relationship with a certain industry, product or services (Khan, 2012). There are two dimensions to loyalty; behavioural and cognitive. Behavioural include the intention of a customer to repurchase, to change and give mainly to the industry, while cognitive is the preference over an industry, advocating for the industry and ability to pay more for services (Taylor, 2007). Enhanced customer loyalty has a lot to do with industry profitability.

2.2 Customer Retention

Customer retention is the continuous maintenance of a trading relationship with customers over a longer period of time. It is the total number of customers dealing with an industry at the end of the financial year in respect to those who were active at the beginning of the year (Dawkins and Reichheld, 1990). Customer retention is very important for the existence of an industry because it drives profitability. Industry is expected to increase retention activities mostly to all loyal customers. Traditional product-oriented industry aims at customer retention as major objectives by putting in more effort to be a more customer-focused element in line with the principles of marketing (Kapai and Moronge, 2015). When an industry retains only 5% of customers, such industry's profit gear up by 25% to 125% (Alshuredi & Alkurdi, 2012). Larivière and Van den Poel (2005) believes turning customer relationship to be more personal has a lot to do with customer retention.

2.3 Related Work

Different efforts have been put in place in building an effective prediction model for customer retention using different techniques. For a better understanding of how several studies have built their prediction models, this article reviews few as shown in table 1.

Table 1: Related literature about customer retention in telecommunication industry

Author	Data set	Predictive Method
Kapai and Moronge (2015)	Airtel Kenya	Linear Regression
Adebiyi, Oyatoye and Mojekwu (2015)	Subscriber Database	Markov Chain (Time-homogenous model)
Sharma and Panigraha (2011)	UCI Repository of Machine Learning Database	Neural Network
Su, Coopre, Probinhei and Jordan (2015)	Customer Analytic record	Clustering (K-means)
Flaherty and Heavin (2015)	Publishing Company	Design science - Predictive Decision Support System (PDSS) 1) survival analysis (Cox's regression), (2) logistic regression and (3) decision Trees (Chi-squared Automatic Interaction Detector)

Kapai and Moronge (2015) applied linear regression in a mobile communication industry in order to explore how customer satisfaction, customer care service, sales promotion and tariff structure influences customer retention.

Adebiyi, Oyatoye and Mojekwu (2015) built a prediction model for Nigeria's telecommunicating industry by using Markov Chains and a time-homogenous model to predict customer churn and retention rates.

From the review of the literature, it can be concluded that any of the predictive models can predict customer retention in the different industry. This paper proposes a logistic regression approach to predict customer retention in the telecommunication industry.

3. THE PROPOSED SSLR MODEL

The main objective of industry offering services should be customer satisfaction which in one way or the other bring them back, but if the customer is loyal, he/she will surely come back (Kapai and Moronge, 2015). SSLR Model is proposed based on the outcome of the variable performance called customer retention. Customer satisfaction is a vital element for those in the services industry (Lewis, 2004). The SSLR model has four qualities namely "S" for Service Quality, "S" for Customer Satisfaction, "L" for Customer Loyalty and "R" for Customer Retention.

Service Quality is used in measuring billing issues, value added services issues and Short Message Service (SMS) issues, out of various numbers of consumer complaints. Customer Satisfaction measures the industry performance which occurs when the quality of service exceed customer's expectation, with this formula:

$$CS = SQ > CE$$

Where CS is customer satisfaction, SQ is quality of service and CE is customer expectation. This shows that customer satisfaction can be influenced by two different management

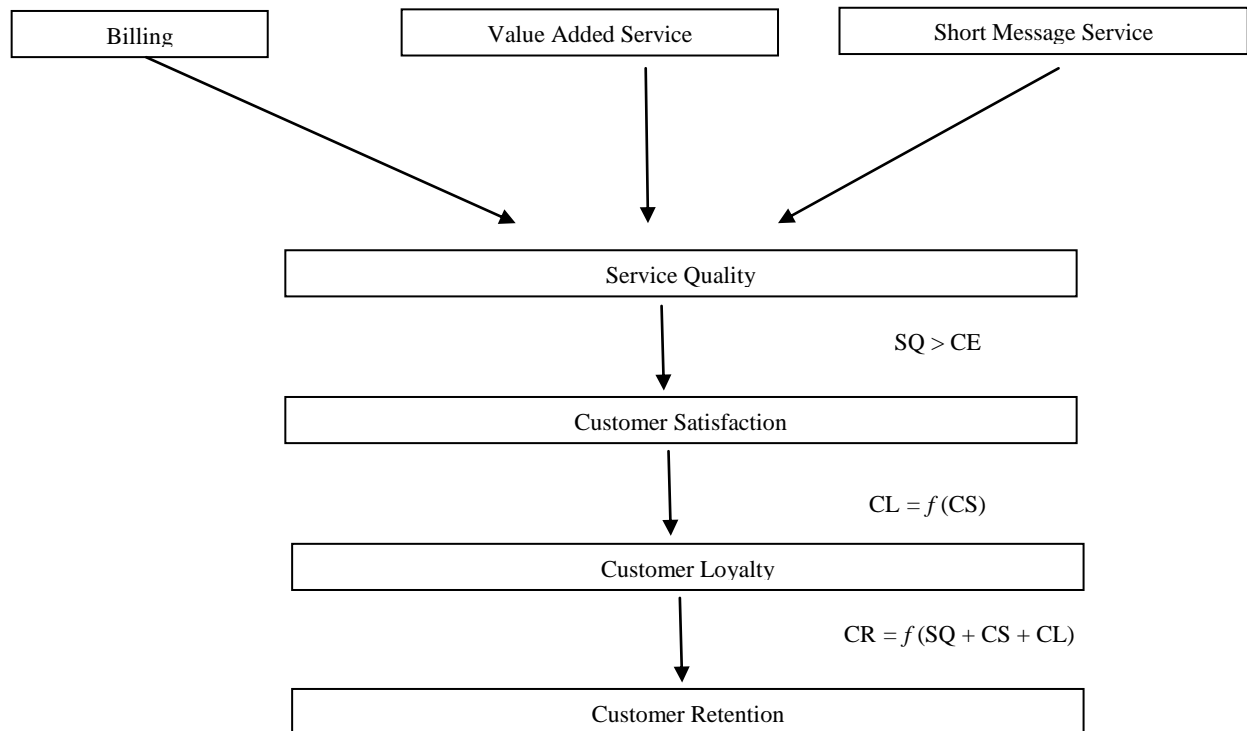


Figure 1: Proposed Model for Customer Retention

(i) customer expectation management and (ii) industry performance management.

Loyalty is used to measure customer satisfaction that is customer loyalty is a function of customer satisfaction, mathematically:

$$CL = f(CS)$$

Where CL is customer loyalty and CS is customer satisfaction.

Retention indicates customer retention as the outcome of the industry performance against three variable namely Service Quality (from the perspective of billing, value-added service and SMS), Customer Satisfaction and Customer Loyalty.

$$CR = f(SQ + CS + CL)$$

Where CR is customer retention, SQ is service quality, CS is customer satisfaction and CL is customer loyalty.

3.1 Comparison of Predictive Model

Predictive Analytics is the most common data mining technique that uses mode to predict what might happens in a particular incident. Alan (2009) described predictive modelling as an aspect of statistics and data mining technique used in getting information from data and using such information to predict events. It comprises of a number of predictors, variables that possibly will assist in future events.

In the use of predictive modelling, the collection of data is the first approach, followed by the formulation of a statistical model, after which prediction is made and the model is revised as more data show-up. According to Nishchol and Sanjay (2012), the steps involved in predictive models are the definition of a project, analysis of source data (exploration), preparation of data, the building of a model, application of model result (deployment) and management of model.

The Predictive model exists in different shapes and sizes and there are more than enough that can be used to create a model. A review of the strength and weakness of the most common types of predictive model is presented in table 2:

Table 2: Comparison of Predictive Model

Types of Predictive Model	Strength	Weakness
Linear Model	It is easy to understand and explain. It can also be adjusted to prevent overfitting. Used in benchmarking, serves as a standard for other models	It creates accurate predictions if and only if the relationship in the data is linear, though it can be overcome.
Decision Trees (Classification and Regression)	Ability to segment a population into smaller and smaller segments. Easy to understand and explain to a non-technical audience.	Most algorithms for developing decision trees are not very effective in using data.
Neural Networks	Ability to take into consideration non-linear characteristics	It is complex. Performance in

	in a data.	usage is substandard to performance base on the acquired data.
Support Vector Machine	It can be used for non-linear decision making.	It is very hard to understand and explain without resorting to the needed mathematics.
Cluster Models	Previous knowledge about behaviour is not necessary.	Struggles with large quantities of data both in terms of the outcome data and predictor variables.

Source: Finlay (2014)

3.2 Dataset

This paper has used the consumer complaint from the Nigerian Communication Commission statistics and reports. The consumer complaints dataset deals with telephony service provider's customer and the data apt to the quality of service they receive. Consumers have the choice of telephony service provider or companies providing them with a telephony service. When these consumers receive good quality of service, they are said to be satisfied thus increase loyalty and then retained which in the long run increase profit. The telecommunication industry concerned here has used three complaints with the highest frequency out of the various listed unresolved complaints by telecommunication consumer and generated a list of precise records related to the study. The logistic regression is implemented on SPSS, the dataset contains three variables worth of information about 18,711 customers, along with the indication of whether or not the consumer is retained or ported. The variables are billing issue, value-added services issue and short message service issue. The data was streamlined to 202 customers based on the incoming and outgoing porting data for the third quarter 2017.

3.3 Logistic Regression

Regression is the measure of the functional relationship between two or more variables in terms of the original units of the data. There are different types of regression namely, simple linear regression, multiple linear regression. Among many types of regression, the most common in data mining is logistic regression which is popularly called a logit model is used to model categorical outcome variable (dependent) where the log odds of the outcome (dependent variable) is modelled as a linear combination of the predictor variable (independent variable). Each predictor is given a coefficient 'b' which measures its independent contribution to variations in the dependent variable which can only take one of the two values: 0 or 1. Logistic regression was used because having a dichotomous outcome variable violates the assumption of linearity in the normal regression. There are three types of logistic regression namely:

- Binary logistic regression model: used to model a binary response
- Ordinal (ordered) logistic regression model: used to model an ordered response

- Nominal (unordered) logistic regression model: used to model a multilevel response with no order.

This paper adopts the use of a binomial logistic regression which predicts the chances that an observation will fall into one of two categories of a categorical dependent variable with respect to one or more continuous independent variables. In this case, the binomial logistic regression was used to understand whether customer retention can be predicted based on billing, value-added service and short message service issues. Where customer retention is the dependent variable measured on categorical scale "yes" or "no" and three independent variables: "billing issue", "value added service issue" and "short message service issue"

4. RESULT AND DISCUSSION

The analysis part of the generated model displays information about the logical regression.

Table 3 shows the model summary in order to depicts how much variation is in the dependent variable.

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	55.293 ^a	.667	.893
a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.			

The Nagelkerke R Square is 0.893 which indicates there is a strong relationship of 89.3% between the predictors (billing, value-added service and short message service issues) and the prediction (customer retention).

Table 4 represents the category prediction that estimates the probability of an event.

Classification Table ^a					
	Observed		Predicted		
			Customer Retention		Percentage Correct
			No	Yes	
Step 1	Customer Retention	No	112	0	100.0
		Yes	9	81	90.0
	Overall Percentage				95.5
a. The cut value is .500					

In the study, 90% were correctly classified for customer retention and 100% for the no customer retention and overall 95.5% were correctly classified. This is a considerable improvement on the 100% correct classification with the constant model so it indicates that the model with predictors is a significantly better model.

Table 5 shows the contribution of each independent variable to the model and its statistical significance.

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Billing_Issue(1)	23.210	4465.886	.000	1	.996	12026312891.738	.000	.
	VAS_Issue(1)	19.195	12710.132	.000	1	.999	217004068.237	.000	.
	SMS_Issue(1)	.000	14411.936	.000	1	1.000	1.000	.000	.
	Constant	-21.203	6793.852	.000	1	.998	.000		
a. Variable(s) entered on step 1: Billing_Issue, VAS_Issue, SMS_Issue.									

In this case, it was noted that billing, value-added service and short message service issues contributed significantly to the prediction (p = 0.996, 0.999 and 1.000 respectively.)

5. CONCLUSION AND FUTURE WORK

A logistic regression was performed to ascertain the effects of billing, value-added service and short message service issues on the likelihood that customers will be retained. The logistic regression model was statistically significant. The model explained 89.3% (Nagelkerke R Square) of the variance in customer retention and correctly classified 95.5% of issues. When billing issues are resolved it is more likely to retain customer while value-added service and short message service issue are associated with the likelihood of exhibiting customer retention. Customer retention prediction is very important in telephone services providers in developing countries. So as to be competitive in the telecommunication industry, telephone service providers are expected to be able to predict customer retention through their complaint, therefore assist in taking proactive steps in retaining the valuable loyal consumers. Therefore to building an efficient model for customer retention is an issue for those both in the industry and academic world. This article suggests that data mining technique can be a bright answer for customer retention and increased loyalty. The final model summary is this article indicated that the model gives more than 95.5% accuracy for the prediction of customer retention.

For future work, other consumer complaint issues can be considered, along with other prediction techniques such neural networks and two prediction techniques can be combined to develop a hybrid model. Finally, the present methodology of customer retention can be tested for other industry like airline, insurance and banking.

6. ACKNOWLEDGMENTS

The authors would like to thank Mr. Akinsola J.E.T for his effort in making this paper work a reality. Also appreciate the anonymous reviewers for their contributions in improving this paper work.

7. REFERENCES

- [1] Adebisi S.O, Oyatoye E.O & Mojekwu E.N (2015). *Predicting Customer Churn and Retention Rates in Nigeria's Mobile Telecommunication Industry Using Markov Chain Modelling*, *Acta Univ. Sapientiae, Economics and Business*, 3 (2015) 67–80, DOI: 10.1515/auseb-2015-0004
- [2] Alan M. (2009). "Introduction to Forecasting Methods for Actuaries." *Forecasting & Futurism Newsletter*, September, pp 6-9. <https://soa.org/Library/Newsletters/Forecasting-Futurism/2009/September/ffn-2009-iss1.pdf>
- [3] Alshuredi & Alkurdi, (2012) in Kapai R. and Moronge, M. (2015). Determinants of effective customer retention in the mobile communication industry in Kenya. *The strategic journal of business and change management*. Vol. 2, Iss. 2 (108), pp 1621 – 1672, Nov 2, 2015.
- [4] Boohene, R. & Agyapong, G.K.Q. (2011). Analysis of the antecedents of customer loyalty of telecommunication industry in Ghana: The case of Vodafone (Ghana). *International Business Research*, 4(1), 229-240.
- [5] Brian O'Flaherty & Ciara Heavin (2015). Positioning predictive analytics for customer retention, *Journal of Decision Systems*, 24:1, 3-18, DOI: 10.1080/12460125.2015.994353
- [6] Dawkins, P.M. and Reichheld, F.F. (1990). Customer retention as a competitive weapon, *Directors & Board*, Summer, pp. 42–47.
- [7] Finlay S. (2014). *Predictive Analytics, Data Mining and Big Data*. Palgrave Macmillan, UK DOI:10.1057/9781137379283
- [8] Heskett, J.L; Jones, T.E; Loveman, G.W; Sasser, W.E & Schlesinger, L.A (2008). *Putting the Service Quality Chain to Work*, Harvard Business Review, July-August
- [9] <https://www.ncc.gov.ng/stakeholder/statistics-reports/consumer-complaints>
- [10] Kapai R. and Moronge, M. (2015). Determinants of effective customer retention in the mobile communication industry in Kenya. *The strategic journal of business and change management*. Vol. 2, Iss. 2 (108), pp 1621 – 1672, Nov 2, 2015.
- [11] Khan, (2012). Impact of Customer Satisfaction and Customers Retention on Customer Loyalty. *International Journal of Scientific and Technology Research*, 1(2), 106-110.
- [12] Larivière, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29(2), 472-484.
- [13] Lewis, M. (2004). The influence of loyalty programs and short-term promotions on customer retention. *Journal of Marketing Research*, 41(3), 281-292.
- [14] Nishchol M and Sanjay S. (2012). Predictive Analytics: A Survey, Trends, Applications, Opportunities & Challenges. *International Journal of Computer Science and Information Technologies*, Vol. 3 (3) , 2012, 4434-4438
- [15] Pogol, G. (2007). *Tips for Cost-Effective Customer Retention Management*, retrieved from www.crm2day.com/library/docs/50577-0.pdf.
- [16] Sharma A. & Panigrahi K. (2011). A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services. *International Journal of Computer Applications (0975 – 8887) Volume 27– No. 11, August 2011*
- [17] Su J., Cooper K., Robinson T., and Jordan B. (2015). *Customer Retention Predictive Modeling in HealthCare Insurance Industry*. BlueCross BlueShield of Florida, Jacksonville, FL.
- [18] Taylor S., Hunter G., and Longfellow T. (2006). Testing on expanded attitude model of goal directed behavior in a loyal context. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behaviour*. Vol. 19 p 18-33