ISSN(Online) : *2456-8910*
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 3, Issue 8, DOI: 10.29027/IJIRASE.v3.i8.2020, 510-513, February 2020

Customer Lifetime Value Prediction: A Study on Multiple Brands Purchase of Consumer Packaged Goods

Arundhati Ghosh, Srinath Naidu
Department of Computer Science & Engineering, Amrita School of Engineering, Bengaluru,
Amrita Vishwa Vidyapeetham, India
arundhatigr@gmail.com, r_srinath@blr.amrita.edu

***Abstract*--Customer is a valuable asset for any business. For the development of any business, business must be interested to understand the intangible value lying within the customer which can be capitalized for the long-term revenue generation. The measurement of customer lifetime value in different periods helps to understand the main factors in the business growth. The objective of the study is to find methodology that can be used to predict customer lifetime value for individual customer of business. The study provides right way to think about customer lifetime value. In this study we deployed customer lifetime value prediction system on CPG (Consumer packaged goods) companies those manufacture one category of personal care product for their customers. We proposed a way to observe underlying behavioural characteristics of customers in different time period that can generate different values of variables. Finally, we concluded with a CLTV (Customer Lifetime Value) model, capable of predicting CLTV (Customer Lifetime Value) for customers who made multiple brands purchase of CPG (Consumer packaged goods).**

Keywords
Consumer Packaged Goods (CPG), Customer Lifetime Value (CLTV), Pareto/NBD (Negative Binomial Distribution), RFM (Recency, Frequency, Monetary).

I. INTRODUCTION

Business is an organization or economic system where goods and services are exchanged for one another or for money. A business has two components namely: (a) Owner and (b) customers. Any business starts with some investment. Objective of any business is to sell the output to enough number of customers in order to make profit. Expectation is to generate profit in a consistent manner.
The year 1980, marked as turning point of wealth creation. Management realized that the principal asset of a company was in fact its intangible assets like brand, copyright, patent, market structure, company reputation etc. The old concept that was used to measure the value of a company in terms of its building, land, plant and machinery is changed. The belief is now that the real value of business lies outside the business itself i.e. in the mind of its potential customers.
CLTV (Customer Lifetime Value) prediction is the study of finding set of high value (low value as well) customers for a business. This is considered as an important problem in commerce, because its vast scope. We can do lot more analysis with the potential customers and come up with new business strategies. The study of CLTV (Customer Lifetime Value) can be extended in that way.
Data used for this study contains customers transaction data. Dataset contains one-year transaction data. Data are provided by a company that conducts surveys in a region of India. The dataset covers most of the well-known brands within the product category. In this study, we considered two major brands mostly used by the customers during the time period of one year.
In the first part of implementation we deploy existing model known as RFM model proposed by Fader, Hardie. We analyse the result. Then try to improve CLTV (Customer Lifetime Value) formula by introducing different latent behavioural variables of the customers. To verify model’s performance, we prepare second questionnaire to collect data set by conducting another survey on the same population in the same specified area. We deploy our CLTV (Customer Lifetime Value) model on that data set. The objective is to determine how the modified CLTV (Customer Lifetime Value) model can be used for better prediction.
In this study customers made mixed purchases. We considered only two brands having highest frequency (number of times customer made purchase).
Therefore, our main contributions are: Developing model to find CLTV (Customer Lifetime Value) under noncontractual business setting, CPG domain where customers made mixed purchase and incorporation of latent variables (that captures customer purchasing behaviour) to modify standard CLTV (Customer Lifetime Value) formula (on single brand purchase) with the expectation of better prediction.

II. RELATED WORK

A lot of statistical models are studied so far to determine customer purchasing behaviour. Simple parametric (probability)model like Negative Binomial Distribution can be used. Models for noncontractual setting are (1) BG/BB (Beta Geometric)/ (Beta Binomial) model (2) The Pareto/NBD (Negative Binomial Distribution) model (3) The BG/NBD (Beta Geometric)/ (Negative Binomial Distribution) model [13]. All models are used for non-contractual business settings where a customer is free to select any product at any time.

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 3, Issue 8, DOI: 10.29027/IJIRASE.v3.i8.2020, 510-513, February 2020

Pareto/NBD (Negative Binomial Distribution) describes the rate at which customers make purchases and the rate at which they churn. The model uses three parameters. The variables required to estimate Pareto/NBD (Negative Binomial Distribution) model is less. To capture customers purchasing behavior, information is collected by following customer-by-customer. We need only three pieces of information for every person: interval between time of last purchase and present (recency), the number of transactions that a customer made (frequency), and the total time horizon for which data are collected [12].

The BG/NBD (Beta Geometric)/ (Negative Binomial Distribution) model, like the Pareto/NBD (Negative Binomial Distribution) model, is also used for non-contractual business settings where a customer is free to select any product at any time.

It describes the rate at which customers make purchases and the rate at which they churn. The model uses three parameters. Recency, Frequency and Monetary value (RFM).

The BG/BB (Beta Geometric)/ (Beta Binomial) model is also used for non-contractual business settings and it works like Pareto/NBD (Negative Binomial Distribution) model.

Finally, Fader, Hardy proposed a model that connects Recency, Frequency, Monetary (RFM) model with CLTV (Customer Lifetime Value) by using iso-value curves in the year 2005 [12]. It is a three-dimensional plot. The plot is used to visualize the interactions and trade-offs among RFM (Recency, Frequency, Monetary) and CLTV (Customer Lifetime Value). They also proposed scoring model for CLTV (Customer Lifetime Value).

However, these models showed satisfactory results for few case studies where they were applied, found difficult to explain customers purchasing behavior by using limited number of variables in their respective CLTV (Customer Lifetime Value) model [3]. In this study, we made an attempt to incorporate variables that capture customer purchasing behavior and where they made multiple brand purchase. We used machine learning approach to achieve this goal.

III. METHODOLOGY

The present study is both explorative and empirical in nature. The explorative part examines different published papers. We also studied different case studies in the field of noncontractual business setting and CPG (Consumer Packaged Good) domain.

In the context of empirical study, we conducted a pilot survey with a sample of 25 consumers to examine the feasibility and appropriateness of the questionnaires and rectify it. We took a random sample of 150 customers. As the customers in and around in a particular zone in India are aware of various brands of CPG (Consumer packaged goods). Different steps of this empirical study are: (1) Preliminary identification of the factors influencing brand valuation. (2) Prepare questionnaires with precise and specific sets of questions with the intension of capturing customer purchasing behaviour as much as possible. (3) Conducting a pilot survey with a sample of 25 consumers to examine the feasibility and appropriateness of the questionnaire and rectify it. (4) Collect a sample of size 150. (5) Identification of the variables that contribute more on customer purchasing behaviour. (6) Computation of CLTV (Customer Lifetime Value) model. (7) Application of our CLTV (Customer Lifetime Value) model on collected data. (8) Prepare new questionnaire by appending new questions that focus more on addressing the causes for brand switching. (9) Conduct a survey with the new questionnaire on the same population of same sample size. (10) Verification of model performance.

I. IMPLEMENTATION

A. *Variable Selection*

Our model incorporates a set of variables to predict customer lifetime values of customers. Every customer of one of the categories of CPG (Consumer packaged goods) was assigned with a value known as customer lifetime value at the end of deployment of the CLTV (Customer Lifetime Value) scoring model. The model computes CLTV for the customer for brand j and is represented as $CLTV_{ij.}$ Table 1 shows variables used in the model.

TABLE I

LIST OF VARIABLES

Variable Names	Meaning
H	Time Horizon (1 year)
t	Time Zone (6 months)
ω_{ij}	Spend at time period t for customer i of brand j
μ_{ij}	Revenue for brand j
$\psi_{ij}{}'$	Spend at time period t for customer i of brand j'
$\pi_{ij}{}'$	Revenue for brand j'
f_1	Frequency of purchase of brand j by customer i
f_2	Frequency of purchase of brand j' by customer i
d_j	Discount offered by brand j
$d_{j'}$	Discount offered by brand j'

ISSN(Online) : 2456-8910
International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE)
Volume 3, Issue 8, DOI: 10.29027/IJIRASE.v3.i8.2020, 510-513, February 2020

B. System Architecture

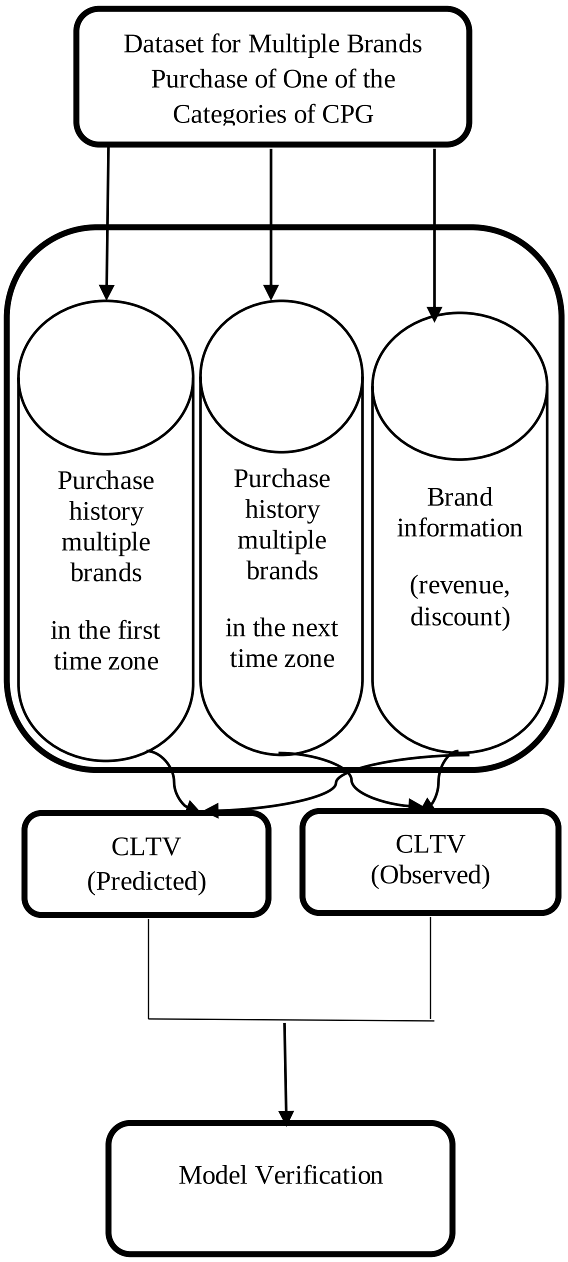

Fig. 1. System Architecture. The solid arrows represent flow of data. Customer data are collected and stored into the database according to the information they are carrying. Stored data are used for CLTV computation.

Customer Lifetime Value Model

Our CLTV (Customer Lifetime Value) model is inspired by standard formula proposed by Fader, Hardie in the year 2013.

Standard formula for CLTV (Customer Lifetime Value) is

$$\text{CLTV} = \sum_{t=0}^{T} m \frac{r^t}{(1+d)^t} \qquad (1)$$

where m = net cash flow per period (if alive)

r = retention rate
d = discount rate
T = horizon for calculation

Standard formula predicts CLTV for ith customer for jth brand. This is suitable for those case studies where customers made single brand purchase. Other possibility like when customers made multiple brand purchase along with the brand j, i.e. j', is not captured. Although the model takes into account retention rate of customers during the entire time horizon.

We proposed another CLTV model by incorporating other variables from TABLE I using training dataset. In this model our objective was to capture those cases where customers made multiple brand purchase.

Proposed CLTV model

$$\text{CLTV}_{ij} = \sum_{t=0}^{H} \frac{\omega_{ij}\mu_{ij}}{(1+d_j)^{t/H}} - K\frac{\psi_{ij'}\pi_{ij'}}{(1+d_{j'})^{t/H}} \qquad (2)$$

Where,

$$K = \frac{f2}{f1+f2}$$

Model performance was verified by comparing RMSE (Root Mean square Error) value obtained from standard formula.

V. RESULT

To verify model performance, we computed RMSE value from of the models (1) and (2). The RMSE values obtained are 34.51 and 24.89 respectively. The study is based on multiple brands purchase made by customers. Hence single purchase of j' brand by a customer in the next time zone has more impact on CLTV for jth brand. What-if analysis can be performed to improve model performance. TABLE II shows comparison between standard formula and the new CLTV formula.

TABLE II

RMSE FOR EXISTING AND NEW CLTV MODELS

Model	RMSE
Standard Formula	34.51
Proposed Model	24.89

RMSE (Root Mean Square Error) values shows new model (2) performs better compared to standard formula (1) for CLTV (Customer Lifetime Value).

VI. CONCLUSION

We described and deployed our CLTV (Customer Lifetime Value) model for multiple brands purchase of one of the categories of CPG (consumer packaged goods). In the first half of the study we describe literature and also discussed about research gaps and challenges. Second half we concentrated on variable selection that can better explain customer purchasing behavior. We prepared questionnaire accordingly. In the next part of the study we described system architecture which in turn explains work flow. Finally, we built our CLTV (Customer Lifetime Value) model and later on model performance was verified.

REFERENCES

[1] "An Introduction to Probability Models for Marketing Research Peter S, Fader University of Pennsylvania, Bruce G.S Hardie London Business School 26th Annual Advanced Research Techniques Forum June 14-17,2015.

[2] "Creating an RFM Summary Using Excel " Peter S. Fader Bruce G. S. Hardie December 2008.

[3] Chen YL, Kuo MH, Wu SY, Tang K (2009). Discovering recent, frequency, and monetary (RFM) sequential patterns from customers purchasing data. Electron Commerce Res. Appl., 8: 241-251.

[4] Hosseini SMS, Maleki A, Gholarmian MR (2010). Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty. Expert Syst. Appl., 37: 5259-5264.

[5] Yeh IC, Yang KJ, Ting TM (2008). Knowledge discovery on RFM Model using Bernoulli sequence. Expert Syst. Appl., 36: 5866-5871.

[6] Ricardo Baeza-yates, Di Jiang, and Beverly Harrison. 2015. Predicting .e Next App .at You Are Going To Use. Proceedings of the 8th ACM International Conference on Web Search and Data Mining (2015).

[7] Oren Barkan and Noam Koenigstein.2016. Item2 Vec: Neural Item Embedding for Collaborative Filtering. Arxiv (2016), 1–8.

[8] Albert C. Bemmaor and Nicolas Glady.2012. Modeling Purchasing Behavior with Sudden" Death": A Flexible Customer Lifetime Model. Management Science 58, 5 (5 2012) 1012–1021.

[9] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. 2001.. e Elements of Statistical Learning. Vol. 1. Springer Series in Statistics, Springer, Berlin.

[10] Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model Fader. et.al 2005.

[11] Probability Models for Customer Base Analysis Fader, Hardie 2013.

[12] RFM and CLV: Using iso-value curves for customer base analysis Fader, Hardie 2004.

[13] Customer-Base Analysis in a Discrete Time Noncontractual Setting Analysis Fader, Hardie 2010.