Customer Segmentation based on Lifetime Value

Nikita Naidu and Surya Putchala

Abstract—Businesses are built around Customers. Understanding customer value is by far the most important thing that impacts its viability and sustainability. Customer lifetime value(CLV) gives an understanding of how profitable a customer will be throughout his journey with the business. Therefore, modelling CLV becomes one of the most critical and challenging problem. In this paper, we have described some of the ways of calculating customer value ranging from Historic CLV to predictive CLV in different business settings: Contractual and Non-Contractual. Further, we present the results of our experiments on CLV modelling in a non-contractual business settings using Pareto/NBD probabilistic modelling technique. We describe our method of classifying the customers into *gold*, *silver* and *bronze* classes according to their Lifetime Value (LTV). We suggest the methodology of using *Soft Margin* for improving classification accuracy. This improved classification accuracy of the dataset under study to 74% which is encouraging.

Index Terms—Customer Lifetime Value(CLV)/ Lifetime value (LTV), Customer Segmentation, Historic CLV, Predictive CLV, Pareto/NBD probabilsitic model.

1 Introduction

There are many definitions for customer value but the most popular definition is the amount of future revenue generated by the customer. Businesses can easily segment their customers into low, medium and high value segments and target their resources to the right group maximizing the returns using CLV information. If the customer value is understood, we can:

- Determine which customers to invest in
- Identify new customers and markets to target
- Agree which product and service lines should be offered and promoted
- Change pricing to extract more value
- Identify the unprofitable customers
- Understand where to cut costs and investments that are not generating growth

A robust understanding of the lifetime value of the customers can provide a clearer view of the valuation of a business, in addition to the potential opportunities to increase the value. Customer lifetime value (CLV) is an important metric to understand the customers and attribute the value the customer brings to the business. Customers differ in their purchasing habits in terms of frequency of visits, recency of visits and the amount spent in their visits. Hence, the customers differ in the value they make for the business. The Pareto principle states that 80% of the revenue is generated by the 20% of the customers. CLV helps business understand their most profitable users and tailor their marketing strategies targeting their most profitable customers to maximize the return on investments.

There are many techniques available in literature on modelling CLV. CLV can be historic or predictive. Historic CLV can be simply calculated from user's past behaviour without any estimation on his future behaviour. Predictive CLV, however, requires modelling of customer's purchase rate to estimate how frequently he will buy in the future and customer's average life span. We have described these in detail in section 2. In [1] Peter and Bruce have explained how businesses can be broadly classified into contractual and non-contractual type and presented brilliant methods for modelling predictive CLV. We have selected Pareto/NBD modelling as an appropriate technique for the e-commerce retail data under study. The results and data description has been provided in section 3.

2 Calculating Customer Lifetime Value

Historical CLV is calculated based on the past transactions of the customer and does not involve any estimate on the future transactions. On the other hand, Predictive CLV models the purchase behaviour of the customer and makes prediction on the future transactions to estimate the potential revenue that can be generated by the customer.

2.1 Historical Customer lifetime value

Average Revenue Per User (ARPU) is one of the simplest way of calculating historic CLV.

ARPU : Average Revenue Per User

This method is very easy and most used method. It involves calculating average revenue of every customer per month and then multiplying by 12 to get total CLV for a year.

Customer	Purchase Date	Amount
James	Sep 07, 2016	$190
James	Oct 10, 2016	$100
Mary	Jan 01, 2017	$50
Mary	Jan 02, 2017	$75
Mary	Feb 20, 2017	$150

TABLE 1: Customer purchase details

Consider a toy example showing customer purchses in table 1. Suppose, we want to calculate aggregate CLV of

	Avg. Revenue per month	ARPU
James	(\$190 + \$100)/ 6 = \$48.33	(\$48.3 + \$137.5)/ 2 = \$92.9
Mary	(\$50 + \$75 + \$150)/ 2 = \$137.5	

TABLE 2: Aggregated CLV using ARPU

the customers as on date 01 March, 2017. We can see that customer James has bought two times in past 6 months and customer Mary has bought 3 times in past 2 months.
Their ARPU can be calculated using formula:

$$\frac{Total\ purchase\ amount}{Total\ months} \tag{1}$$

(Refer table 2 for ARPU calculations). Individual level ARPU per month(using formula (1)) can be aggregated to obtain the ARPU of the entire population. ARPU = \$92.9. To obtain a 6 month or 12 month CLV multiply ARPU by 6 or 12.

ARPU model can be improved by considering recency of visit as a factor to distinguish between an active or inactive customers. It may also include frequency of visits as a measure to give more weightage to frequent customers. However, there are big limitations in ARPU model. It does not capture heterogeneity in customer behaviour. It considers all the customers the same. Further, ARPU model assumes that the customer behaviour will be constant throughout the customer journey which is not true. ARPU estimations could often be misleading and give a poor estimate of CLV.

2.2 Predictive Customer Lifetime Value

This method takes customer dropouts/churn into consideration to determine average lifespan of the customer with the organisation. However, it is not trivial to predict the future purchase pattern for every customer as it depends on the type of business. There are two types of B2C businesses: Subscription based services (such as Cable TV, Digital Newspapers, Mobile Plans, Club memberships, Software) and Open Retail (Groceries, Electronic, food stores etc.,). In a subscription service, end of customer journey can be definitely known when the customer subscription is up for renewal. These services are driven by *Contractual setting* throughout the Customer Journey. The Customer/Business has to overtly and wilfully stop the service. The purchases can also happen in a regular or irregular intervals. On the contrary, in a retail stores or e-commerce set up *(Non-contractual)*, there is no event that marks the end of customer journey. Thus, the latent parameters i.e. customer lifespan, purchase rate and monetary spent has to be derived from the deviations in customer behaviour. A customer can leave any time or even come back afterwards for a purchase. There is no event that marks customer churn. Hence, these different business scenarios will have to be modelled considering different features and metrics in order to estimate customer churn and thus, CLV.

Predictive CLV in Contractual setting:
In contractual settings, the customer CLV can be calculated by divinding the population into cohorts and capturing the customer retention ratio dynamics. CLV can be calculated using the following standard formula:

$$CLV = \sum_{t=0}^{T} m \ \frac{r^t}{1+d^t} \tag{2}$$

where,
m = net cash flow per period (if active)
r = retention rate
d = discount rate
T = horizon for calculation

Consider following illustration of a hypothetical scenario in a contractual setting.
Assume:
- Each contract is annual, starting on January 1 and expiring at 11:59pm on December 31.
- An average net cashflow of \$200/year.
- A 10% discount rate

Table 3 shows the cohorts by acquisition date and table 4 shows year wise retention ratio in each cohort.

What is the expected residual value of the customer base at December 31, 2016?

Aggregate retention can be calculated using:

$$\frac{2504\ +\ 3264\ +\ 4367\ +\ 6334}{3264\ +\ 5179\ +\ 7339\ +\ 10000} = 0.64$$

Expected residual value of the customer base at December 31, 2016 using standard formula (2):

$$\$26469\ \times \sum_{t=1}^{T} \$200\ \frac{0.64^t}{1\ +\ 0.1^t}$$

(T can be any number of years)

	2012	2013	2014	2015	2016
Cohort 1	10000	8559	6899	3264	2504
Cohort 2		10000	6334	5179	3264
Cohort 3			10000	7339	4367
Cohort 4				10000	6334
Cohort 5					10000
	10000	18559	23233	25782	26469

TABLE 3: Number of active customers year by year in each cohort

	2012	2013	2014	2015	2016
Cohort 1		0.86	0.81	0.47	0.77
Cohort 2			0.63	0.82	0.63
Cohort 3				0.73	0.60
Cohort 4					0.63
Cohort 5					-
		0.86	0.71	0.68	0.64

TABLE 4: Retention ratio in each cohort

Predictive CLV in Non-Contractual setting:
There are several different proababilistic models but they all evolve around the same assumptions and modelling framework.

A typical probabilistic modelling framework:

1 These models estimate the latent (unobserved) parameters of the contractual setting x: Customer Lifespan, Purchase Rate and Monetory Spent; by treating behaviour as if it were random (probabilistic).
2 Selecting the distribution that can fit this behaviour x denoted as $f(x/\theta)$. These are individual latent traits.
3 Specifying a distribution that can characterize population level distribution of the latent trait θ denoted as $f(\theta)$

In the next section, we have described one of the most popular probablistic modelling framework: Pareto/ NBD for non-contractual setting:

Pareto - Negative binomial distribution (Pareto/ NBD):
This is one of the most popular probabilistic modelling technique for calculating CLV in non-contractual setting. It makes following assumptions on the latent parameters :

1 The individual level purchase rate λ is characterized by poisson distribution.
2 λ is distributed across the population by gamma distribution characterised by r and α.
3 The individual level churn rate μ is characterized by Exponential gamma.
4 μ is ditributed across the population by gamma distribution characterised by s and β

The model basically takes recency, frequency, total transactions in the calibration period and total observed time for each customer in the population. It ouputs the four parameters r, α, s, β using which the purchase rate and the dropout/churn rate can be estimated. Thus, the expectation of the purchase rate $E(\lambda) = \frac{r}{\alpha}$ and expectation of dropout rate $E(\mu) = \frac{s}{\beta}$.

Estimating CLV using Pareto/NBD model:
The model outputs expected number of transactions in the future at customer level. It takes individual customers recency, frequency, total transactions and observation time to produce expected number of transactions T at the end of period t. CLV can then be simply calculated using:

$$CLV = T \times AOV \tag{3}$$

where,
T = Number of transactions in time t
AOV = Average Order Value

Thus, there are many ways of modelling CLV. The performance of these models must be validated by comparing predicted CLV and actual customer generated revenue in the future. In the next section, we present our method of modelling CLV for a retail domain and show how it can be used to segment the customers.

3 Experimentation and Results

We modelled predictive CLV using Pareto/NBD model to obtain lifetime value for every customer in the dataset. The dataset was divided into calibration data and holdout data. The customers in calibration data were then segmented as gold(70% total revenue), silver(20%) and bronze(10%) based on the percentage of predicted total revenue(by CLV) generated by them. We compared the segmentation results to the actual revenue generated by the customers in holdout data. We also came up with a distinctive approach of segmentation : *Soft Margin*.The classes (gold, silver, bronze) are only logical segments generated using filters (revenue cutoff). The adjacent classes(gold and silver, silver and bronze) are seperated with many customers falling on the margin or very close to the margin. Hence, instead of a hard-margin we propose a soft-margin with ± 5% of total revenue on either side of the margin. Therefore, our new segments are gold(70 ± 5%), silver(20% ± 5%), bronze(10% ± 5%). The idea is to classify the people falling in the overlapping region as members of both classes. Meaning, people generating revenue between (65% ± 75%) are both gold and silver customers. Similarly, people generating revenue between (15% ± 5%) are both silver and bronze customers.

Dataset and Experimental setup :
We have used Online Retail dataset from UCI ML repository [3]. The data consists of 541909 instances. Preprocessing includes removing of negative transactions(return, order cancellations), Invoices with Order Value 0 and duplicate line items. We divided the data into calibration and holdout set as follows:

- Dataset length : Transactions between 2010-12-01 to 2011-12-31 (13 months)
- Cut-off date : 2011-05-31
- Calibration set : All transactions with Invoice date $\geqslant$ cut-off date
- Holdout set : All transactions with Invoice date $\leqslant$ cut-off date
- Total customers in entire dataset : 5213
- Customers in Calibration set (Old/Observed) : 3131
- Customers in Holdout set (New/Unobserved) : 2082

Modelling and Parameter Estimation:
The model was built on the calibration set and the CLV for all the cutomers (New and Old) was predicted for the next 7 months (holdout set duration). Customers were observed for only 6 months and based on their purchase patterns our model learned the churn rate and purchase rate of the customers as shown in table 5. The expectation of the Purchase rate (transaction rate) of the general population is 0.32 transactions per month. However, the dropout rate based on the inter purchase transaction time is very low (0.0000617). The distributions of both are shown in fig 1 and fig 2.

r	α	s	β	purchase rate E(λ)	churn rate E(μ)
0.58	1.81	6.17e-05	426.22	0.32	1.449e-07

TABLE 5: Model parameters

Validation and Performance:
The model fits the calibration set well with log likelihood -6575.19. As shown in fig 3, the model's estimation of frequency of repeat transactions for the customers in the calbiration period closely matches their actual frequency.

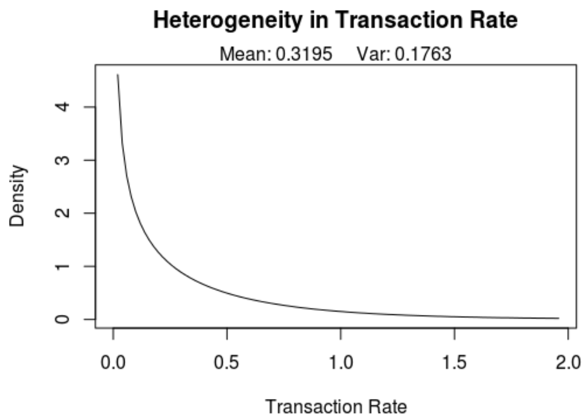

Fig. 1: Distribution of transaction rate

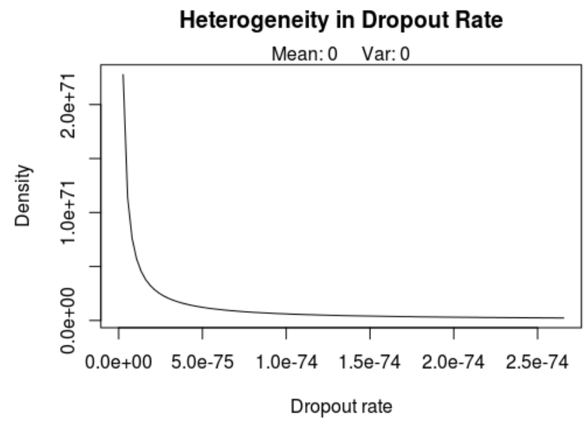

Fig. 2: Distribution of dropout rate

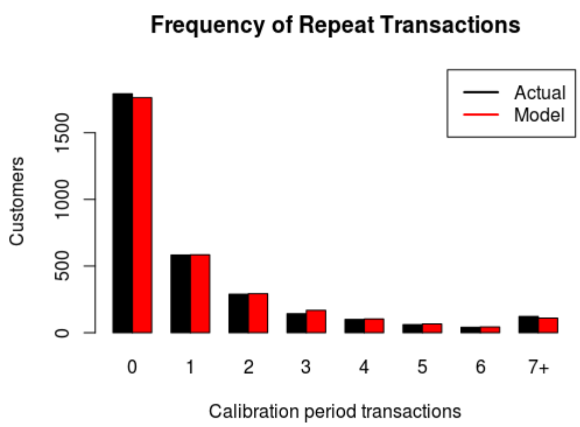

Fig. 3: Actual Vs Predicted Frequency of transactions in calibration period

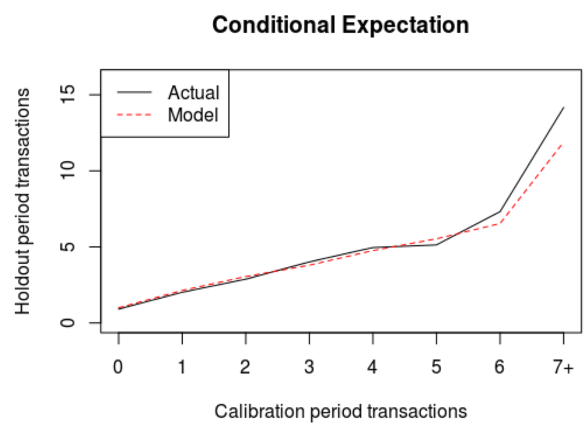

Fig. 4: Actual Vs Predicted transactions in calibration and holdout period

Similarly, the model's conditional expectation on the number of transactions in the future given the number of transctions in the calibration period for a customer closely follows the actual purchase pattern (fig 4). Thus, we can conclude that model has learnt the purchase rate and churn rate very well.

Segmentation of customers and Results:
Customer lifetime value has been calcuated using formula (3)

$$CLV = T \times AOV$$

where,
T = Customer transactions in time t
AOV = Average Order Value

The customers have been segmented into Gold(G), Silver(S) and Bronze(B) class based on the percentage revenue predicted by CLV. We grouped the customers in two ways: Regular margin and Soft margin.

Regular margin :
Gold (G) - Customers generating 70% of the total revenue
Silver (S) - Customers generating remaining 20% of the total revenue
Bronze (B) - Customers generating remaining 10% of the total revenue

Soft margin :
Gold (G) - Customers generating 65% of the total revenue
Gold Silver (GS) - Customers generating remaing 35% - 25% of the total revenue
Silver (S) - Customers generating remaining 25% - 15% of the total revenue
Bronze (BS) - Customers generating remaining 15% - 5% of the total revenue
Bronze (B) - Customers generating remaining 5% of the total revenue

Further, we calculated the actual revenue generated by the customers in the holdout period and segmented them to the G, S, B buckets. Table 6-8 shows the confusion matrix obtained with regular margins for the different classes of customers. We obtained prediction accuracy of 0.64 for all the customers (refer table 9). Classification using Soft margin improves the prediction accuracy to 0.74.(Refer table13). The confusion matrix are shown in table 10-12. Any point in segment GS is classified as both Gold and Silver(same for BS). Thus, improvement in prediction accuracy looks obvious as some of the bands have overlapping classes. But, the highlight is the improvement in Sensitivity and Specificity

		actual		
		G	S	B
	G	535	323	234
predicted	S	142	527	874
	B	28	257	2293

TABLE 6: Confusion Matrix for Regular Margin All Customers

		actual		
		G	S	B
	G	303	145	189
predicted	S	97	297	411
	B	26	205	1458

TABLE 7: Confusion Matrix for Regular Margin Old Customers

		actual		
		G	S	B
	G	232	178	45
predicted	S	45	230	463
	B	2	52	835

TABLE 8: Confusion Matrix for Regular Margin New Customers

	Gold		Silver		Bronze		
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Prediction Accuracy
All	0.76	0.88	0.48	0.75	0.67	0.84	0.64
Old	0.71	0.88	0.46	0.8	0.71	0.78	0.66
New	0.83	0.88	0.5	0.69	0.62	0.93	0.62

TABLE 9: Performance measure for each class of customers with Regular Margin

		actual		
		G	S	B
	G	586	246	186
predicted	S	91	604	563
	B	28	257	2652

TABLE 10: Confusion Matrix for Soft Margin All Customers

		actual		
		G	S	B
	G	339	95	189
predicted	S	61	347	230
	B	26	205	1639

TABLE 11: Confusion Matrix for Soft Margin Old Customers

		actual		
		G	S	B
	G	247	151	45
predicted	S	30	257	285
	B	2	52	1013

TABLE 12: Confusion Matrix for Soft Margin New Customers

	Gold		Silver		Bronze		
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	Prediction Accuracy
All	0.83	0.90	0.55	0.84	0.78	0.84	0.74
Old	0.80	0.90	0.54	0.88	0.80	0.78	0.74
New	0.89	0.90	0.56	0.81	0.75	0.93	0.73

TABLE 13: Performance measure for each class of customers with Soft Margin

in each class. (Sensitivity and specificity are calcuated in the standard way of calculating True Positive Rate and True Negative Rate.)

As per pareto principle, we expect a smaller group of people contributing to very large section of revenue. Here, the gold customers are essentially expected to generate high revenue (70%). So, to validate the CLV values we have plotted each class (B G S as predicted by CLV in the holdout period) against the actual amount of revenue generated by them in holdout period. We see that indeed 17% of the population segmented as G class by the model is actually generating 63.3% revenue. Refer pie chart below (fig.5)

Thus, our model predictions are highly accurate and can be used to yield very useful insights on customer value.

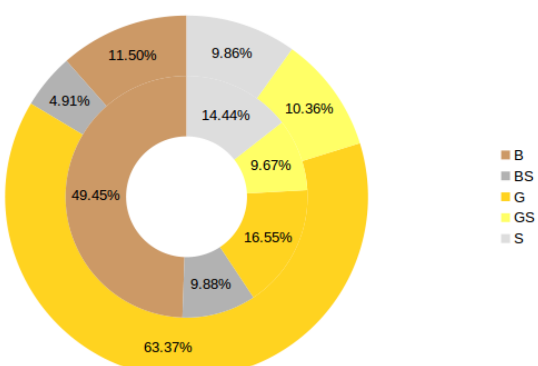

Fig. 5: Predicted CLV bins versus Actual revenue

4 Future Work

As described by Peter and Bruce in [2], Customer's Average transaction value (average order value) is an imperfect estimation of his mean transaction value. The paper also presents gamma gamma extension to the Pareto/NBD model for estimating the mean transaction value for a customer. Our future work involves extension of our model to estimate the mean transaction value using gamma-gamma model extension and study its effect on Customer Lifetime Value.

5 Conclusion

The calculation of CLV is non-trivial and depends on the type of Business - Contractual or non-contractual business. CLV helps simplifying allocation of marketing budget to various channels. It also helps in formulating strategic plans for acquiring new customers, engage and retain existing customers and increase loyalty to create a sustainable revenue stream. It thus becomes critical to predict CLV with high accuracy. Our model can predict the CLV with 74% accuracy for both Observed and New customers. Further, the model predicts the *gold* class customers with 0.83 sensitivity and 0.90 specificity; successfully identifying the most important class of customers.

References

[1] Probability Models for Customer-Base Analysis *Peter S. Fader, Bruce G. S. Hardie* http://www.brucehardie.com/talks/art05_cba_tutorial_handout.pdf
[2] The Gamma-Gamma Model of Monetary Value *Peter S. Fader, Bruce G. S. Hardie* http://www.brucehardie.com/notes/025/gamma_gamma.pdf
[3] Online Retail, UCI ML repository https://archive.ics.uci.edu/ml/datasets/Online+Retail

Nikita Naidu is a Data Scientist at Cappius Technologies with area of expertise in Artificial Intelligence, Machine Learning and Statistical Analysis. She is currently working on designing and modelling of Customer Insights framework which is a platform to understand the customers in various businesses and delivers actionable insights useful in all phases of customer journey - Retention, Engagement and Acquisition. Earlier she has worked on Root Cause Analysis and Continous Improvement of business with Accentue Services Pvt Ltd. Nikita has pursued Bachelor of Engineering, CS from University of Pune and Master of Technology, AI from University of Hyderabad and has been awarded graduation Gold Medal for academic excellence.

Surya Putchala provided thought leading consulting solutions in the areas of Business Intelligence, Data Warehousing, Data Management and Analytics to Fortune 500 Clients over the last two decades. Surya has a tremendous zeal to create Analytics products that bring significant improvements in Business Performance. Currently the vision behind the comprehensive customer experience management platform called ”Capptix”. He is passionate about areas related to Data Science, Big Data, High Performance Cluster Computing and Algorithms. He constantly endeavours to evangelize the adoption of quantitative techniques for decision making in various verticals. He held senior leadership roles with firms such as GE capital, Cognizant, Accenture and HCL. He graduated from IIT Kharagpur.