*International Journal of Research in Engineering and Science (IJRES)*
*ISSN (Online): 2320-9364, ISSN (Print): 2320-9356*
*www.ijres.org Volume 9 Issue 7 ǁ 2021 ǁ PP. 47-50*

# Finding the Most Profitable customer using CLV

Prathima J, Vaishnavi M, Perumalraja R and Kamalesh S
*Department of Information Technology*
*Velammal College of Engineering and Technology,Madurai,TamilNadu*

***Abstract—*** *In marketing, Customer Lifetime Value (CLV) is a foreseeing of the net profit contributed to the whole future relationship with a customer. CLV represents the total amount of money a customer is expected to spend in your business. Hence, CLV is an important metric to find the most profitable customer. Traditional techniques, like Recency, Frequency and Monetary Value (RFM) for segregation of profitable or Non-Profitable customers, based on their history of purchases. In a competitive business world, it is necessary to maximize profit by analyzing customer behavior and to identify and target customers with the greatest potential CLV value over time. Hence, various models such as, Markov chain model, RFM model, weighted RFM mode, CART and Regression model, Pareto/NBD Model,etc., were studied by the researchers to calculate CLV. We predicted CVL more accurately with stability for even large and short time interval dataset by using BG/NBD model and Gamma-Gamma model. We used BG/NBD to find the expected number of transactions and Gamma-Gamma model to find the Revenue per transaction, thereby CLV value found with high accuracy. We also discussed our approach with real time e-commerce online retailer (Amazon) for finding the CLV.*

***Keywords—****CLV; RFM; BG/NBD model and Gamma-Gamma model*

---

Date of Submission: 29-06-2021 Date of acceptance: 13-07-2021

---

## I. INTRODUCTION

Customer lifetime value (CLV), represents, the net profit attributed to the entire future relationship between the company and the client. CLV is a measurement of how valuable a customer is to a company within a period. It can be used to determine which customer segments are the most profitable (high value) and which are not (low value) [1] by predicting how much profit will the customer contribute to the company in the future.

There have been other measures as well which are fairly good indicators of customer loyalty like Recency, Frequency and Monetary Value (RFM)[2], a statistical model based on the Pareto principle.

- **Recency**: The item purchased by the client very recently.
- **Frequency:** How many times the client has purchased the items in their time
- **Monetary Value:** How much amount the client has spent.

The customers who satisfied with the RFM means, which customers have more recently and have a high frequency and total monetary contribution are said to be the best customers in this approach. However, it is possible that the most profitable customer of today may not be the same as tomorrow. customers who can be good at a certain point of the period and may not be good later and a bad customer turning to good by gaining. Being active in the future is a very important part of the calculation of the probability of a customer in CLV calculation, which is used to calculate customer loyalty. whether a customer will continue his relationship with it in the future or not, it is very important for a firm to know. CLV helps to understand the customer's behavior in the future and thus enables them to allocate their resources according to their Behavior.

CLV is defined as the present value of all future profits obtained from a customer over his or her entire lifetime of relationship with an organization. A very basic mathematical formula to calculate the CLV of a customer is

$$CLV_i = \Sigma^{T}_{t=1} \frac{(\text{Future contribution margin})_{it} - (\text{Future Cost})_t}{(1+S)^t}$$

where,

- t denotes time index,
- S denotes the discount rate
- T denotes the number of time periods considered for estimating CLV,
- i denotes customer index,

To calculate the customer lifetime value, there is a formula based on theoretically are,

**CLV = (Expected number of transaction) * (Revenue per transaction)**
In this approach, the BG/NBD and Gamma-Gamma models are used to calculate the CLV.BG/NBD model is used to find the Expected number of transactions. Gamma-Gamma model is used to find the Revenue per transaction
Using CLV, you can better understand the different people among your customers, we effectively target or personalization the customers.

- CLV Focusing on Long-Term Company-Wide Growth
- CLV becomes a competitive market for e-commerce companies in 2021
- CLV is a customer-centric metric, and a powerful base to retain valuable customers, increase revenue from less valuable customers, and improve the customer experience overall.

The following pages of the paper are organized as follows: Section II discusses the literature review. Section III discusses the proposed model. Section IV discusses the implementation of our project and the results of our project are discussed in the next section. Section VI discusses the real-time in E-Commerce site. Section VII is the conclusion.

## II. LITERATURE REVIEW

The existing CLV calculation methods are classified into two kinds: the models that take into account past customer behavior and the models that consider both past and future behaviors. Every past customer behavior group model has unique parameters which are directly related to the model's characteristics. Among the models, RFM is the most widely used one and it has been utilized in marketing areas.

Rust, et al. (2000), have proposed the Markov chain model, which is a deterministic model, where it involves randomness and uncertainty values, which is only based on previous events [3]. This model is inappropriate for over sufficiently short time intervals, Hence Fader, Hardie, Lee. (2005) has proposed the RFM model is based on three quantitative metrics: Recency; Frequency; Monetary. This model is not predicted as a precise quantitative model and not suitable for large datasets. To overcome this, Liu, Shih. (2005),[4] has proposed the WRFM model used Analytic Hierarchy Process, AHP, to determine the relative weights of RFM parameters. Since an influential parameter is a weight, Therefore this model is called weighted RFM or WRFM but it is Complex and not suitable for large datasets. Due to this lag I-Cheng et al. (2009), have proposed the RFMTC model (Recency, Frequency, Monetary Value, Time, Churn rate) an augmented RFM model, that calculates the probability of a customer buying at the next promotional or marketing campaign. The cons of the RFMTC model are done based on only first purchase and Probability. CJ Cheng. (2012), has proposed the Markov chain model and neural networks are used to find the possible purchasing frequency of a customer, while neural network approaches are used to estimate the profit. The cons of this model are if the time interval is too short, then Markov models are inappropriate because they are not random individual displacements, but rather are deterministically related in time. This suggests that Markov models are inappropriate over sufficiently short time intervals.

Wu, lin, Liu (2014) has proposed the CART (Classification And Regression Tree), A Classification And Regression Tree (CART), is a predictive model, which explains how an outcome variable's values can be predicted based on other values. A CART output is a decision tree, this model is Unstable, Not suitable for large datasets, High variance, Overfitting leads to wrong predictions which are the main lag in this model to overcome from this Glady, N., Lemmens, A. & Croux. (2014), has proposed the Pareto /NBD(negative binomial distribution) Model and Gamma -gamma sub-model Pareto/NBD [6] is another type of model used to predict the future activity of customers and uses previous data as the primary input. It will not segregate active and inactive customers and the Gamma-Gamma sub-model, there is no relationship between the monetary value and the purchase frequency, Hence Lemmens and croux (2015) and Sunder, Kumar, Zhao. (2016) have proposed the Basic structural model and structural model are Relies heavily on across and within variation in customer purchases. Arsie P. Mauricio, John Michael M. Payawal, Maida A. Dela Cueva, Venusmar C. Quevedo. (2018), has proposed the Logistic Regression and MLP(Multilayer perceptron), is a predictive model for one or more dependent target variable based on a previous variable, The cons of this model is defined the relationship only between dependent binary variables. It uses many parameters which result in redundancy and inefficiently.

## III. THE PROPOSED MODEL

In this approach, the BG/NBD model and Gamma-Gamma models are used to calculate the CLV.

### A. BG/NBD model

In this research, we used to predict the expected number of transactions by using the Beta Geometric/Negative binomial distribution model which was proposed by Fader, Hardie, and Lee. It is a good model for RFM type of problems by modeling discrete-time of data and comparing the forecast result with the actual data. Since CLV calculation is completely based on RFM type, BG/NBD model suits this research [3].

- There are three steps to implement this model, they are getting ready with parameters
- Creating a sales forecast using the parameters
- Predict the future purchase of a customer base of parameters estimated and past behavior analysis.

To implement this the three needed parameters are Recency (When his last transaction occurred), Frequency (number of transactions in a specific period of time), and monetary (the amount the customer spends over the same period of time). These three parameters are represented as

$$(X = x, tx, T) \qquad (1)$$

Here x is a number of transactions over a period (0, T] and tx is a time at the last transaction where $0< tx <= T$. The next step is to make some assumptions based on some mathematical distributions to predict the expected number of transactions [9],

- **Poisson distribution** with rate lambda describes the number of transactions of an active customer in a period of time
- **Gamma distribution** with shape parameter r and scale parameter a describes Heterogeneity or deviation in customer behavior or transaction
- **Geometric distribution** describes the dropout point when Users become inactive after some transaction
- **Beta distribution** with shape parameters alpha and beta describes Heterogeneity in dropout probability
- Transaction rate and dropout probability vary independently across users.

The model fits the distribution to the historic data and learns the distribution parameter and then uses them to predict future transactions of a customer.

### B. Gamma-Gamma model

The Gamma-Gamma Model can predict the most likely value per transaction in the future. The Gamma-Gamma model is used to model the monetary value.

The model is based on the following three general assumptions:

- The monetary value (e.g., $, £, e) of a customer's given transaction varies randomly around their average transaction value.
- Average transaction values vary across customers but do not vary over time for any given individual.
- The distribution of average transaction values across customers is independent of the transaction process.

To calculate the average number of the transaction [10]

$Z = \Sigma\ xi=1\ Zi/x$

Where,

- zi = Distributed normal.
- X= Number of observations.
- Zi/x = observed mean transaction value

The frequency (xi) and monetary value (zi) data for each individual (i = 1, .., I), the gamma distribution function is

$$E(M \mid p, q, \gamma, mx, x) = \frac{(\gamma+mxx)\ p}{(px+q-1)}$$

$$= \frac{(q-1)\ \gamma p + px * mx}{(px+q-1)\ q-1 \qquad px1}$$

- p is shape
- v is scale parameters of gamma distribution for transactions Zi,
- q is shape and
- γ is scale parameters for gamma distribution of v (p is constant by assumption individual-level coefficient of variation is the same for all customers).

## IV. IMPLEMENTATION

We have collected the dataset from UK-based and registered non-store online retail during 2010 -2011. By these datasets, we have developed a web application to calculate the CLV value corresponding to the uploaded CSV. Files that consist of transaction details. The internal working of an application is done using BG/NBD model and Gamma-Gamma model developed using Jupyter Notebook anaconda software in python language.

### A. Dataset

The transaction data used to develop this model is transactions that occurred between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail. Characteristics of a dataset are multivariate and sequential data. It contains eight attributes such as Invoice No, Stock Code, Description, Quantity, Invoice Date, Unit Price, Unit price. Numeric, Product price, Customer ID, and Country. This data contains almost 541909 instances of transaction.

**Table**. 1.Sample data

| Voice No | Stock Code | Description | Qty | Invoice Date | Unit Price | Customer ID | Country |
|---|---|---|---|---|---|---|---|
| 536365 | 85123A | T-LIGHT HOLDER | 6 | 01-12-2010 | 2.55 | 17850 | UK |
| 536365 | 71053 | WHITE METAL LANTERN | 6 | 01-12-2010 | 3.39 | 17850 | UK |
| 536365 | 84029G | CREAM CUPID HEARTS COAT HANGER | 8 | 01-12-2010 | 2.75 | 17850 | UK |
| 536365 | 84029E | KNITTED UNION FLAG HOT WATER BOTTLE | 6 | 01-12-2010 | 3.39 | 17850 | UK |

Already Daqing Chen, Sai Liang Sain, and Kun Guo [5] proposed a model based on Recency, Frequency, and Monetary to segmented into various categories of customers in the business using k-means clustering algorithm and decision tree induction.The main characteristics of the consumers in each segment have been identified clearly. But now we are creating a customer lifetime value prediction model to find the most profitable customer

### B.GUI output

We have developed the CLV system by using Jupyter notebook anaconda software with the programming language python. Then we used an anvil uplink to develop a user interface for an application and to connect the developed model with the user interface and come up with a complete software system. Anvil uplink takes a machine-learning model in a Jupyter notebook, and turns it into a web application.

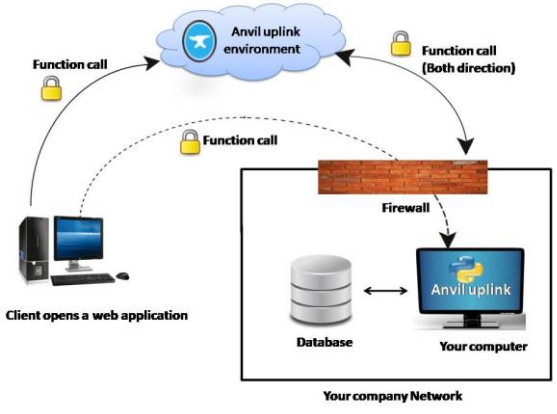


**Fig.** 1.Architecture of GUI

***B. Implementation of a model***

These complex models are done by a simple python library function called lifetime package. This package is primarily used to build customer lifetime value calculation models. BG/NBD model is available as **Beta-Geo-Fitter** class in Lifetimes package. First thing is that to fit transaction data into beta-Geo-Fitter

The next thing is to find that whether a customer is now active or not, the probability of being active is calculated based on the recency and frequency of a customer this is done with the lifetime package. Methods from the lifetime package are used to implement the model are

- **model. conditional probability alive():** This method separates the currently active customer by computing the probability of a customer with a historical transaction.
- **Plot_probability _alive_matrix(model):** This function plots and helps to visually analyze the relationship between recency & frequency and the customer is active.

This method finds the active customer based on the below two key points.

If a customer has bought many times (frequency) and the interval between the first & last transaction is high (recency), then his/her probability of being active is high. If a customer has less frequency (bought once or twice) and the interval between the first & last transaction is low (recency), then his/her probability of being active is high. We consider customers who made repeat purchases with the company i.e., frequency > 0. if frequency = 0, it means there are only one-time customer and are considered as inactive. The next step is to train the model to predict the expected number of transaction

- **model.conditional_expected_number_of_purchases_ up_to_time():** Calculate the expected number of transactions up to time t for an individual from the history of the dataset given.

Thus the BG/NBD model is used to predict the expected number of future transactions

Next, the monetary aspect of the CLV calculation that is revenue generated per transaction is found by Gamma-Gamma Model. First, need to find the relationship between frequency and monetary value. If the correlation seems very weak. Hence, we can say that our assumptions are satisfied and we can fit our data into the model by using gammagammaFitter. Next, we can predict the expected revenue for each transaction and Customer Lifetime Value using the model.

- **model.conditional_expected_average_profit():** the average profit per transaction for a group is computed by conditional expectation

Thus the gamma-gamma model is used to predict the revenue per future transactions. The final step is to calculate the clv value by just making a product of the expected number of transactions and revenue generated per transaction for the time period of t. It is also found by the method called customer_lifetime_value()

**model.customer_lifetime_value():** This method computes the CLV for one or more customers. This method takes two parameters as input i.e., expected number of transactions from BG/NBD model and revenue generated per transaction by gamma-gamma model for the period of time t.

Thus customer lifetime value is calculated to find the most profitable customer by using both BG/NBD and Gamma-Gamma models. The values we have calculated for CLV is not the actual profit but a sales value. To get the net profit for each customer, we can multiply sales value with profit margin. The company can now use this clv value to target customers to promote their product and increase their sales and profit.

## V. RESULT AND DISCUSSION

We developed a Web application to find CLV value for the corresponding Uploaded file. Our application consists of three pages. A home page consists of two-button names: find top profitable customers and find individual CLV value. clicking on this button will navigate you to the corresponding page.


**Fig.** 2. Home page of GUI

The next page is to find the top profitable customer, here it consists of an option to upload csv. File and enter the number of top customers needed, once the find button is clicked the list of top customers will be displayed in GUI. It also consists of a back button to go back to the home page.


**Fig.** 3. Find top profitable customer page of GUI

The next page is to find the CLV value for individual customers. once you uploaded csv. File and enter the customer ID the corresponding customer's CLV value will be displayed in GUI.


**Fig.** 4.Find individual customer CLV page of GUI

To make sure the accuracy of found CLV value we have done research by increasing the dataset from minimum count to reasonably high data count we initially we use only 50,000 datasets while it could not able to predict the expected number of transaction exactly, then we slightly increased the count of dataset until 5 lack the accuracy also increased as dataset increasing finally The Predicted CLV value seems to be accurate. This shows that the number of transaction data used to predict the value of a system is directly proportional to the accuracy of value arrived by the system.

We have also done other research to prove the accuracy of a system. first, we fit transaction data of six months into our model and then predicted the CLV for the next one month, we verified the result manually with the actual dataset with predicted data for the next month. by this experiment, the deviation in the comparison is negligible hence we proved the Predicted CLV seems to be accurate for our model. Thus both BG/NBD and Gamma-gamma model are the best model to predict the customer lifetime value

## VI. CLV CALCULATION IN AMAZON (REAL TIME E-COMMERCE SITE)

CLV is one of the most valuable calculations from your customers and will allow you to make quick refinements and increase your profits in a short duration of time. Success is all about finding the right customers at the right time. Now We are calculating the lifetime value of your current customer base, you'll be able to start crafting that target and win over those customers that really increase your profit.

A good understanding of CLV on Amazon will help you in segmenting your customers. By segmenting you can trigger and push your best customers into buying for the next time. One of the most effective ways to increase CLV is to increase customer satisfaction[8].5% increase in customer satisfaction can increase CLV by 25% to 95%.

The procedure used by the amazon e-commerce website to find CLV value will be as below steps,

1.Pivot the Data and Sort the Pivot Table by Item-Price
2.Total Number of Buyers will be counted and Consolidate the Total Data
3.Calculate and Count the Number of Unique Orders
4.Add Unique Orders to the Report Summary
5.Calculate Average Order Value, Average Order Frequency, and Annual Customer Value
6.Calculate Average Customer Lifetime and Total Customer Lifetime Value

## VII. CONCLUSION

We have calculated the CLV value and thereby found the most profitable customer, by using the RFM approach with BG/NBD model and the Gamma-gamma model with high accuracy. An organization can maximize its profit by analyzing customer behavior and identify and target customers with the greatest potential CLV value over time.

Nowadays, In every digitized shop, every customer's transaction details are generated automatically by using that, the CLV value will be calculated and it makes the company find more profitable customers and increase their revenue. Knowing the importance of CLV will allow you to focus more time and effort on high-value customers, increasing your revenue of an organization, and improving relationships with your long-term customers at the same time.CLV gives you crucial insight into how much money you should be spending on acquiring your customers by telling you how much value they'll bring to your business in the long run.

## REFERENCES


[1]. Arsie P , John Michael M, Maida A and Venusmar C, "Predicting Customer Lifetime Value through Data Mining Technique in a Direct Selling Company". *Proc. of Int'l. Conf. on Industrial Engineering, Management Science and Application*, 2016, pp.1-2.

[2]. Tarun R , "Customer Lifetime Value Measurement using Machine Learning Techniques" ,2011,pp.5-7.

[3]. Esmaeili Gookeh M and Tarokh M J ," Customer Lifetime Value Models: A literature Survey". Proc. Journal of Industrial Engineering & Production Research,2013, Vol 24, Number 4, pp. 317-336.

[4]. Abdulkadir H , Merve S , Halil Ibrahim C and Omer Faruk S, "An Empirical Assessment of Customer Lifetime Value Models within Data Mining",*Proc. of Int'l. Conf. on Forum and Doctoral Consortium*,2018, pp.34-38

[5]. Daqing C, Sai Laing S and Kun G ," Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining",Proc. Database Marketing & Customer Strategy Management ,Vol. 19, pp 197–208.

[6]. Sien C , "Estimating Customer Lifetime Value Using Machine Learning Techniques" ,2018, pp.20-24

[7]. Nicolas G, Aurélie L and Christophe C, "Unveiling the Relationship between the Transaction Timing, Spending and Dropout Behavior of Customers",2012,pp. 5-7.

[8]. Han J , Li Jiying and Fan Wenyan ,"Customer Lifetime Value Model Based on Customer Satisfaction",2010,pp.2-4.

[9]. Peter S. Fader ,Bruce G. S. Hardie and Ka Lok Lee, "Implementing the BG/NBD Model for Customer Base Analysis in Excel",2005,pp.1-2.

[10]. Peter S. Fader A and Bruce G. S. Hardie "The Gamma-Gamma Model of Monetary Value" (2013)pp.1-3

[11]. Lim Chia Y and Vincent KT , "Customer Relationship Management: Computer-Assisted Tools for Customer Lifetime Value Prediction",*Proc. of Int'l. Conf. on Symposium on Information Technology, ITSim* ,2010,pp. 1180-1185.