iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.68233

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

A Comprehensive Review of Customer Lifetime Value by using NBD and Gamma-Gamma Model

Aditya Singh[1], Mr. Ajeet Kumar[2], Shubham Singh[3], Mithilesh Kumar Verma[4], Priyansh Gupta[5]
JSS Academy of Technical Education Noida

Abstract: This project aims to predict and classify customer lifetime value based on historical transaction data of an e-business entity between December 2015 and December 2017. The research begins with exploratory data analysis followed by RFM analysis [2] so as to develop a foundational understanding of the customer's behavior and classify customers into groups based on their purchase behavior. This is through two-step modeling when approximating the CLTV. In this case, while BG/NBD generates a sort of estimate on how that particular customer will make that purchase, the Gamma-Gamma model will approximate the monetary average per head. This establishes that the project provides an in-depth analysis of the customer value gained from the outcome of the two models, identifies high-value customers, and classifies them into four different groups. Such insights are applicable in the implementation of optimal market targeting strategies that facilitate customer retention strategies and enhance revenue maximization. By this, the project manifests the promise that predictive advanced analytics holds in enhancing decision-making processes and furthering business performance.

I. INTRODUCTION

Understanding customer value is critical to the firms seeking sustainable growth. Questions such as: What is the value of a customer to firm? How might their behaviors change over time? Should investments be targeted to specific customers to ensure profitability? Who are the most valuable customers? are critical in the development of effective marketing strategies. One of the most popular techniques used to answer these questions is Customer Lifetime Value (CLV). By analyzing how CLV enables the businesses to build strategies that serve the needs of customers as well. With such a strategy, business satisfies the customer and emerges strong over competitors. When a relationship is not contractual, for example with business models that are not contractual, customers can even be discontinuing with them any moment. In such business-related industries, CLV comes highly valuable. In these settings, correct CLV models allow one to predict customer behavior and pinpoint those who are likely to generate a lot of value in the long run. Such information can be used to allocate the right resources, focus on retaining high-value customers, and minimize the chances of customers leaving. This research project aimed at measuring CLV as a means of helping organizations make informed decisions based on data. We started by doing segmentation by The exploratory data analysis was conducted to split customers on both demographic and behavioral characteristics. After the first segmentation phase, a set of models was applied to the estimation of customer lifetime value. The BG/NBD model was applied to forecast the future transactional frequency within the loyalty programs and, on the other hand, Gamma-Gamma was utilized for estimating the subsequent revenues generated from those transactions. transactions. Further, we enriched our work on customer clustering by conducting K-Means clustering of the CLV to further enrich our outcome. By leveraging these techniques, companies can develop more targeted strategies for customer retention and acquisition, ultimately driving greater profitability and long-term growth.

II. LITERATURE REVIEW

There are many ways to calculate CLV, employing a wide range of scientific methodologies, algorithms, and analytical tools. All these methods are supposed to offer significant insights from history and current business data. Some of the widely used techniques for estimating CLV are probabilistic models, and BTYD models are one of the most fundamental frameworks for such models. There are explicit models developed for the computation of customer lifetime value that include Pareto/NBD and BG/NBD [6] [10].
The decision tree models are another method of calculating CLV. Among the algorithms well known with regard to this are XG Boost, which pertains to tree boosting, and has often been used in classification and regression tasks. For instance, it was utilized on a database of 240,000 passengers of China Eastern Airlines, with over 300 features for every passenger. This approach produced results of the utmost precision because customer economic values are measured from 0 to 100. The next example is a churn prediction model developed using XG Boost that produced a result of 97.9% accuracy for the sample customers of Telkom Indonesia from March 2018, as all of these customers have been in the company for more than 12 months [3].

Decision trees are very widely used for the computation of CLV. For example, a European airline used decision tree models. These models were found more advantageous for the company than using any other strategy such as classification or regression. In addition, an automobile repair and maintenance company in Taiwan employed a decision tree model in the prediction of customer lifetimes. It obtained a maximum accuracy of 78.66% on the training dataset and 77.06% on the testing dataset. The advantage with decision trees is that they are also completely immune to the rigid assumption about the data distributions making them work in good time with both continuous and categorical forms of data [4].

Many companies are increasingly adopting neural network algorithms to improve CLV calculations. For example, an Australian insurance company utilized 12 years of data from 19,174 customers to develop recommendations aimed at maximizing CLV. Once implemented, this approach resulted in higher transaction volumes and increased profitability. These outcomes highlight the effectiveness of both neural networks and decision tree algorithms in calculating CLV, with both methods remaining widely used today [5].

The BG/NBD model is a probability model which is highly used in predicting future behaviors of customers based on past transactions with them. It modifies the base Pareto-NBD model while reducing some of its drawbacks. Developed by Schmittlein, Morrison, and Colombo, the Pareto-NBD model was developed to monitor customer purchase cycles and classify customers as either active or inactive. However, due to challenges in parameter estimation, Dr. Fader, Dr. Hardie, and Mr. Lee introduced the BG/NBD model in 2005 as an improvement [6] [7].

The BG/NBD and Pareto-NBD models differ in how they define customer inactivity. In the Pareto-NBD model, customers are considered "lost" based on external criteria, meaning their inactivity is not directly tied to their last purchase date—they may cease activity at any point after a transaction. Conversely, the BG/NBD model adopts a more internally consistent approach by assuming customers become inactive immediately following their most recent purchase. The BG/NBD model's simplicity and efficiency have contributed to its popularity, as it only requires three specific data points for each customer.

1) Recency (R): Months since the first transaction until the last transaction.
2) Retention (F): The transactions a customer has made to some period of time that help determine the likelihood of buying again.
3) Total Observation Period (T): It is the time period from when the customer had made the first transaction until the last time of observation, indicating the "life cycle" of the customer in the chosen unit of time (e.g., months or years) [7][10].

The BG/NBD model is a valuable tool for businesses to forecast a customer's future behavior, including the frequency of transactions. By analyzing past activities, this model provides an estimate of a customer's potential future value, making it highly effective for calculating Customer Lifetime Value (CLV). For example, in the grocery retail space, BG/NBD model has been successfully applied in order to estimate future counts of transactions that a particular customer would make. BG/NBD model captures purchases made by a customer during fixed periods of time. In actuality, "time period" can be defined as the duration between the first purchase that has been made by the customer and can be in the form of months or years. It helps organizations understand the possible behavior of their clients and, therefore, make better decisions regarding retained customers and the overall business strategies.

The BG/NBD model is quite effective for measuring important customer metrics such as the probability that a customer continues buying, or becomes active again and the number of transactions he might make in the near future based on past behavior. It thus forms a reliable framework to measure the value of customers and predict their future behavior across different companies.

The Gamma-Gamma model focuses on calculating the average monetary value of each transaction. Developed by Dr. Fader and Dr. Hardie, the model aims to determine the monetary worth of individual customers. By leveraging this model, businesses can forecast consumer behavior and estimate the average transaction value during uncertain time periods. This insight supports future trend prediction and informed decision-making. The RFM analysis, often used for tactical purposes, plays a crucial role within the framework of the Gamma-Gamma model.

Key characteristics of the Gamma-Gamma model, as highlighted by Fader, Hardie, and Lee, include:

1) Customer transaction values tend to fluctuate randomly around an average value.
2) Average transaction values evolve over time and vary between customers.
3) The mean transaction values across different customers follow a gamma distribution[6].

Gamma-Gamma model is an important model for any enterprise because it helps rank those customers who do not have frequent transactions but have higher profit margins. The usage of this model will help businesses predict the potential monetary value of transactions and identify those clients who can be best suitable for growth [6][7].

In summary, the combination of the BG/NBD and Gamma-Gamma models offers an effective approach for evaluating customer value.

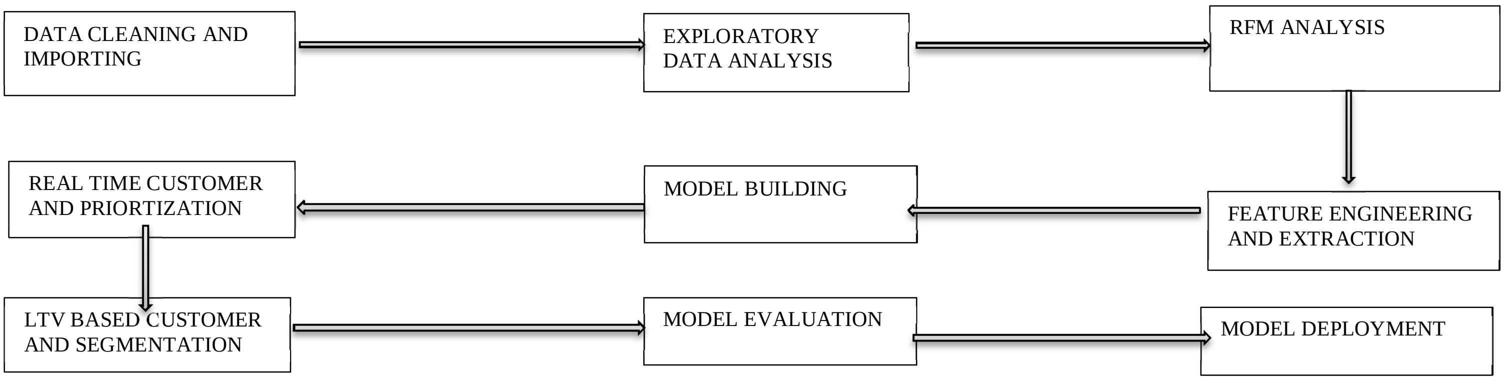
This allows businesses to design targeted promotions for specific customer segments, maximizing their impact and effectiveness. customers works efficiently for organizations. With probabilistic models like BG/NBD, one advantage it comes with is that it may work with very few features and yet deliver more precise results when compared to other alternatives for algorithms such as classification, regression, and decision trees. For example, the number of characteristics that classification algorithms need is 200-300, whereas the BG/NBD model can compute customer lifetime value using only three key characteristics. Also, the BG/NBD model fits well with non-contractual business environments where a customer can make a single purchase and not return again. Considering all these factors, the BG/NBD model has been chosen for this project as it reduces the time taken to perform feature engineering and can also result in more accurate predictions in non-contractual relationships

A. Summary of Research Approaches

The following table summarizes methodologies, datasets, key findings and remarks from significant studies on CLTV.

Title	Methodology	Limitations	Results
M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, "Estimating Customer Lifetime Value Using RFM Analysis of Consumer Purchase Behavior."	Used RFM analysis to estimate CLV.	Limited granularity in segmentation compared to probabilistic models.	Effectively segmented customers into actionable groups for marketing strategies.
N. Ali and O. S. Shabn, "Insights into Customer Lifetime Value (CLV) for Strategic Marketing Success and Its Influence on Financial Performance."	Explored strategic use of CLV insights for marketing and financial planning.	Limited discussion on implementation techniques or specific methodologies.	Showcased the financial benefits of using CLV-driven strategies.
N. Glady, B. Baesens, and C. Croux, "An Improved Approach to Predicting Customer Lifetime Value with a Modified Pareto/NBD Model."	A refined Pareto/NBD model has been proposed to enhance the accuracy of Customer Lifetime Value (CLV) predictions.	Complexity in model adjustments may require extensive validation.	Enhanced prediction accuracy for high-value customer segmentation.
D. Garcia, J. Desirena, G. Desirena, and I. Moreno, "Optimizing Customer Lifetime Value Through Stacked Neural Networks."	Employed stacked neural networks to maximize CLV in the insurance sector.	Requires extensive computational resources and is prone to overfitting with smaller datasets.	Increased transaction levels and profits for the insurance company.
P. Fader, B. Hardie, and K. Lee, "A Simplified Approach to Estimating Customer Lifetime Value: An Alternative to the Pareto/NBD Model."	A modified Pareto/NBD model has been suggested to improve the accuracy of predicting Customer Lifetime Value (CLV).	Simplicity may limit capturing complex customer behaviors.	Simplified CLV calculation with minimal features while maintaining accuracy.
D. C. Schmittlein, D. G. Morrison, and R. Colombo, "Understanding Your Customers: Their Identities and Future Actions."	The Pareto/NBD model was created to forecast customer retention and the frequency of transactions.	Rigid parameter estimation and external determination of customer inactivity.	Established a foundational probabilistic approach for CLV prediction.
H. Casteran, L. Mayer-Waarden, and W. Reinartz, "Modeling Customer Lifetime Value, Retention, and Churn Dynamics."	Combined CLV modeling with retention and churn analysis.	Integration of multiple models increases complexity and computational cost.	Provided comprehensive insights into customer behavior and profitability.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com



Flow chart

III. METHODOLGY

A. Data Loading and Understanding

Excel files containing transactional records are imported into Python using Pandas. Both datasets (2015–2016 and 2016–2017) are combined to form a unified data structure for analysis. Data is then examined for shape, consistency, and missing values. Null or invalid entries in critical columns like Customer ID are flagged and handled appropriately. Duplicate records, if any, are removed to prevent bias in the analysis. These steps ensure a clean and reliable dataset, which is essential for building accurate models [2],[4].

B. Data Preprocessing

Data cleaning involves identifying and treating anomalies such as negative quantities or monetary values. Missing values are either imputed or removed based on their impact on key metrics. Outliers in features like transaction frequency or monetary value are capped using statistical methods such as the interquartile range (IQR). Dates are standardized, and new features like total transaction value (Quantity × Unit Price) are computed. Preprocessing transforms raw data into a structured format suitable for analysis and modelling[4],[6].

C. RFM Analysis and Customer Segmentation

RFM analysis is a technique used to assess customers based on three critical factors: the recency of their last purchase, the frequency of their purchases, and the amount they spend. By assigning scores to these factors, businesses can categorize customers accordingly. For instance, customers with high RFM scores—those who purchase often and spend more—can be targeted with loyalty initiatives or exclusive deals to foster repeat buying. Conversely, customers with lower RFM scores may be the focus of re-engagement strategies, like targeted promotions to encourage them to make additional purchases.

The RFM metrics—Recency, Frequency, and Monetary Value—are used to evaluate customer engagement:

- Recency measures the time elapsed since a customer's last purchase, offering insight into their likelihood of returning [8].
- Frequency gauges how often a customer makes a purchase, reflecting their level of loyalty [7].
- Monetary Value tracks the total amount a customer spends, highlighting their contribution to the business's revenue [6].

Customers are scored on these metrics (e.g., 1 to 5) and grouped into segments such as high-value, medium-value, and low-value tiers. This segmentation helps businesses prioritize customer retention and targeting efforts.

D. CLTV Prediction and Segmentation

The outputs from the BG/NBD and Gamma-Gamma models are integrated to estimate each customer's lifetime value for a specified period (e.g., 3, 6, or 12 months). Based on these predictions, customers are categorized into different tiers, such as high, medium, or low-value. High-value customers are targeted with loyalty initiatives, while low-value customers may be approached through re-engagement campaigns. This segmentation offers valuable insights for creating personalized marketing strategies and optimizing revenue.

IV. EVALUATION METRIC

Models are validated using a holdout dataset and metrics like RMSE (Root Mean Squared Error) to assess the accuracy of predictions. Visual checks, such as comparing predicted versus observed monetary values, ensure model outputs align with real-world trends. The effectiveness of segmentation is evaluated by analyzing intra-segment similarity and business impact, such as revenue growth or churn reduction [4],[9].

V. CONCLUSION

This study attempts to use a hybrid approach of data analysis and predictive modeling techniques in the task of CLTV forecasting. The project started with intense analysis of the transactional data where data quality issues were addressed, buying patterns identified, and RFM analysis conducted. Such groundwork was laid down so that predictive models could be constructed [6],[9].

We applied the BG/NBD model alongside the Gamma-Gamma model to estimate CLTV. The BG/NBD model helped us determine the purchase frequency and the likelihood of customers ceasing to make purchases. Meanwhile, the Gamma-Gamma model assisted in estimating the average revenue per transaction generated by each customer. By combining these models, we were able to accurately predict the future value of each customer and categorize them into various value segments [6],[7].

CLTV and customer segmentation can be achieved accurately and have business implications that are specific in nature. The company will thus know the most valuable customers to target, and the rest of its resources can be channelled in developing more precise marketing techniques. A company can use loyalty programs and customized promotions for its best customers, while strategies in retaining customers who will most likely leave should be formulated. It's a data-driven approach in marketing efficiency while supporting long-term profitability and growth [8],[10].

REFERENCES

[1] L. Vrana, L. Sperkova, M. Kobulsky, P. Jasek, and Z. Smutny, “A Comparative Study of Probabilistic Models for Customer Lifetime Value in Online Retail,” Journal of Business Economics and Management, vol. 20, no. 3, pp. 403–413, 2018.

[2] S. Chen, “Data Mining: Estimating Customer Lifetime Value with Machine Learning Methods,” pp. 26–32, Apr. 2018.

[3] I. Hanif, “Enhancing Customer Churn Prediction Through the XGBoost Classifier,” Media & Digital Department of Telkom, Indonesia, Aug. 2019.

[4] C. J. Cheng, C. B. Cheng, J. Y. Wu, and S. W. Chiu, “Predicting Customer Lifetime Value Using a Markov Chain- Based Data Mining Model: Insights from an Auto Repair and Maintenance Company in Taiwan,” Scientia Iranica, pp. 850–855, Nov. 2011.

[5] D. Garcia, J. Desirena, G. Desirena, and I. Moreno, “Maximizing Customer Lifetime Value Through Stacked Neural Networks: A Case Study in the Insurance Sector,” pp. 3–8, Jul. 2019.

[6] P. Fader, B. Hardie, and K. Lee, “Simplified Methods for Estimating Customer Lifetime Value: An Alternative to the Pareto/NBD Model,” Marketing Science, vol. 24, no. 2, pp. 275–284, 2005.

[7] D. C. Schmittlein, D. G. Morrison, and R. Colombo, “Who Are Your Customers and What Will They Do Next? Insights from Counting Customer Behavior,” Management Science, vol. 33, no. 1, pp. 1–24, 2016.

[8] H. Casteran, L. Mayer-Waarden, and W. Reinartz, “Customer Lifetime Value, Retention, and Churn Modeling,” in Handbook of Market Research, C. Homburg, M. Klarmann, and A. Vomberg, Eds. Springer, Cham, pp. 14–41, Apr. 2017.

[9] M. Mzoughia and M. Limam, “Advancing the BG/NBD Model for Purchasing Behavior Analysis Using the Com-Poisson Distribution,” International Journal of Modeling and Optimization, vol. 4, no. 2, pp. 141–145, 2014.

[10] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, “RFM Analysis of Customer Purchase Behavior: Estimating Customer Lifetime Value - A Case Study,” Procedia Computer Science, vol. 3, pp. 57–63, 2011, doi: 10.1016/j.procs.2010.12.011.