Customer Lifetime Value Prediction

H S Yashaswini[1] and Prabhudeva S[2]
4th Semester MCA, Master of Computer Application[1]
Professor & Director, Master of Computer Application[2]
JNN College of Engineering, Shimoga, India

Abstract: *In this era of Internet, there are many lots of retailers in the e-commerce industry for whom the customers are assets. E-commerce means buying and selling the products through Internet in online mode. E-commerce established many employment opportunities to the people from anywhere because there is no direct interaction between the seller and the buyer. Many people are purchasing things using this e-commerce application. There are many e-commerce websites available for the customers. So the Retailers want to analyze their relationship with the customers so that they can produce or buy the goods according to their requirements. However, this work mainly focuses on predicting the customer lifetime value (CLV) using Beta-Geometric/Negative Binomial Distribution Model (BG\NBD) and Gamma-Gamma Model.*

Keywords: Customer Lifetime Value (CLV), Beta-Geometric/NBD Model, Gamma-Gamma model, Customer Relationship Management (CRM).

I. INTRODUCTION

More people buying things using these e-commerce websites because of quick response. There are many offers available also compared to direct purchase. Here, mainly the retailer has to focus on maintaining customer relationship using Customer Relationship Management (CRM). To maintain this CRM, computation of the lifetime value of the customer which is known as Customer Lifetime Value (CLV) is needed.

As long as a customer remains as a client, retailer's income can be expected from the customer. The measure of this total income for a business is called as Customer Lifetime Value. To measure CLV, certain things had to be known such as, expected number of transactions, revenue per transaction and margin assigned by the company. To simplify this process they need to divide the company's CLV into segments which helps them to focus on high CLV customers and apply the same strategy across the whole customer base. CLV help to understand that, to what extent a customer like the product or the services that the company provides there by the company can understand the area in which they have to improve. Here the complexity arises in acquiring the new customers so it is necessary to retain the existing customers which will help in knowing the high CLV customers to generate more revenue and retaining the existing customers will helps in reducing the expenditure cost of the company. The increasing CLV helps the company to know that, it is making a better impression on the customers.

II. LITERATURE SURVEY

Asiman Mammadzada et al [1], has published a paper in which it is found that the BG/NBD and Gamma-Gamma Model has obtained the satisfactory results by considering the total amount of financial transactions and the number of transactions made by the customers.

Arsie P. Mauricio et al [2], has used Binomial Logistic Regression, Multilayer Perceptron Neural Network and Multiple Linear Regression model for predicting the CLV. The accuracy of Binomial Logistic Regression model was found to be 94.1% for churn and not-churn customers.

Benjamin Paul Chamberlain et al [3], describe the Random Forest and Neural Network model to predict CLV for ASOS global e-commerce company. They have developed a hybrid model that combines the above mentioned models.

Pavel Jasek et al [5], in their paper, it has made comparison on performance of Petro family models for CLV calculation Even though there are no ideal models, BG/NBD and Petro/NBD models comparatively gave the good and stable results.

Than Than Win et al [8], has published the paper, it has implemented Random Forest Model to predict the CLV. The performance of this model was well and it has given good f1-score and high accuracy for high class customers.

Yi Wang et al [10], describes on CLV. Here, Artificial Neural Network (ANN) and Multi-Layer Perceptron (MLP) models were utilized to calculate CLV. Further, it is observed that there is difficulty in calculating CLV for the prepaid customers due to associated cross-selling and up-selling effects.

III. PROPOSED METHODOLOGY

3.1 Algorithm: CLV

- Step 1: Load the dataset.
- Step 2: Removing rows containing null values.
- Step 3: Calculating the total sales.
- Step 4: Calculating frequency, recency, T, monetary_value.
- Step 5: Fitting the BG/NBD model.
- Step 6: Calculating the probability of alive customers and predicting the number of transactions using BG/NBD model.
- Step 7: Fitting Gamma-Gamma model.
- Step 8: Calculating expected average sales.
- Step 9: Calculating Customer Lifetime Value (CLV) in Days, Months and Years.
- Step 10: Classification of customers as Platinum, Gold, Silver and Bronze using CLV.

3.2 Block Diagram

Block diagram of the proposed work is shown in figure1 below. The Online retail dataset used for the proposed work is taken from the UCI Machine Learning Repository. This dataset is cleaned and manipulated based on the requirements of the methodology. Then Total sales, Frequency, Recency, T and Monetary Value will be calculated. Here, the dataset will be trained using BG/NBD and Gamma-Gamma models. Using these two models it will calculate the Probability of the alive customers, Predicting number of transactions, calculates expected average sales, later it will also predict the CLV in days, months and years after it will classify the customers based on CLV.

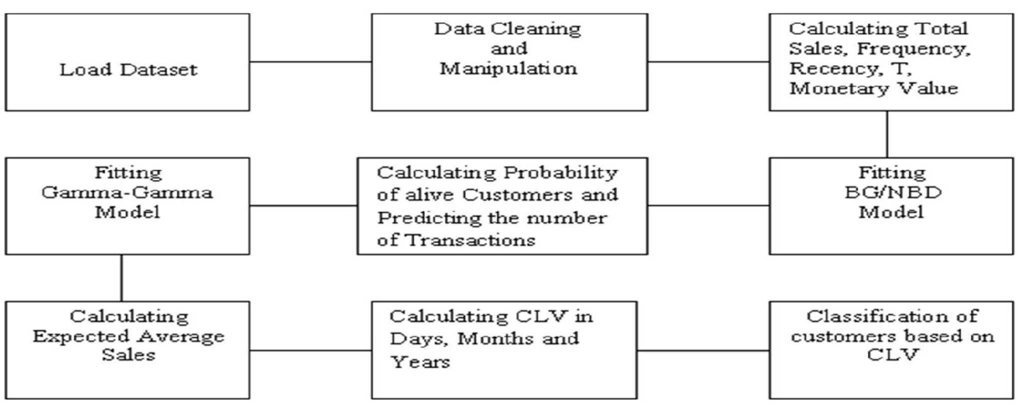

Figure 1: Block Diagram of the proposed work

IV. RESULT ANALYSIS

The result analysis is used to give the experimental results of the proposed work that is calculating the total sales, frequency, recency, T, monetary value, probability of alive customers, predicting number of transactions, expected average sales, CLV.

Sl No.	Customer ID	Invoice Date	Total Sales (Rs. Thousands per day)
1	17850.0	2010-12-01	15.30
2	17850.0	2010-12-01	20.34

3	17850.0	2010-12-01	22.00
4	17850.0	2010-12-01	20.34
5	17850.0	2010-12-01	20.34

Table 1: Calculating the total sales

The above Table 1 represents the calculation of Total sales, which can be obtained by multiplying the quantity with unit price. It will helps for further analysis.

Table 2: Calculating the probability of alive customers and predicting number of transactions

Sl No.	Customer ID	Frequency	Recency	T (age in time units)	Monetary Value	Probability of alive customers	Predicted number of transactions
1	14911.0	131.0	372.0	373.0	1093.661679	0.999966	8.95
2	12748.0	113.0	373.0	373.0	298.360885	0.999971	7.73
3	17841.0	111.0	372.0	373.0	364.452162	0.999962	7.59
4	15311.0	89.0	373.0	373.0	677.729438	0.999964	6.10
5	14606.0	88.0	372.0	373.0	135.890114	0.999955	6.03

In Table 2, it represents the calculation of probability of alive customers and predicting the number of transactions, which is done by using frequency, recency and T.

Table 3: Calculating the expected average sales

Sl No.	Customer ID	Frequency	Recency	T (age in time units)	Monetary Value	Probability of alive customers	Predicted number of transactions	Expected average sales
1	12347.0	6.0	365.0	367.0	599.701667	0.999698	0.47	603.04
2	12348.0	3.0	283.0	358.0	301.480000	0.9999177	0.27	306.73
3	12352.0	6.0	260.0	296.0	368.256667	0.999405	0.56	371.03
4	12356.0	2.0	303.0	325.0	269.905000	0.999478	0.22	277.56
5	12358.0	1.0	149.0	150.0	683.200000	0.999486	0.25	704.70

In Table 3, it represents the calculation of expected average sales, which is done by using frequency and monetary value.

Table 4: Calculating the customer lifetime value

Sl No.	Customer ID	Frequency	Recency	T (age in time units)	Monetary Value	Probability of alive customers	Predicted number of transactions	Expected average sales	Manual predicted CLV	CLV
1	12347.0	6.0	365.0	367.0	599.701667	0.999698	0.47	603.04	283.43	14.17
2	12348.0	3.0	283.0	358.0	301.480000	0.9999177	0.27	306.73	82.82	4.14
3	12352.0	6.0	260.0	296.0	368.256667	0.999405	0.56	371.03	207.78	10.39
4	12356.0	2.0	303.0	325.0	269.905000	0.999478	0.22	277.56	61.06	3.05
5	12358.0	1.0	149.0	150.0	683.200000	0.999486	0.25	704.70	176.18	8.81

The above Table 4 represents the calculation of Customer Lifetime value (CLV), which is obtained by multiplying predicted number of transactions, average number of transactions and margins (it is set by the company).

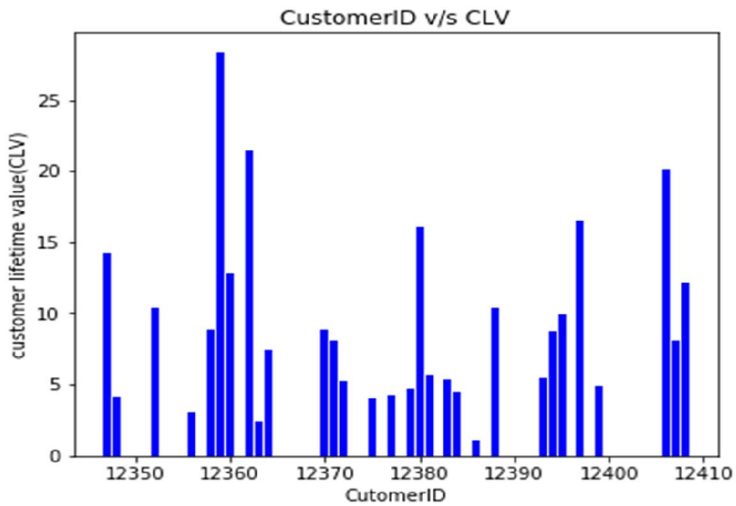

Figure 2: Graph for visualizing the CLV based on Customer ID

In Figure 2, the bar graph represents vizualization of CLV based on CustomerID. In x-axis, it is considering customerid and in y-axis, it is considering CLV. The above graph represents the profit to the company (i.e. CLV) by individual customer.

4.1 Output Samples

In the below snapshots, snapshot A calculates the total sales. Snapshot B will calculate the probability of customers who are alive and number of transactions of the customer. Snapshot C calculates the average sales. Snapshot D calculates the customer lifetime value by multiplying the manual_predicted_clv with the margins assigned by the company.

	CustomerID	InvoiceDate	Total_sales
0	17850.0	2010-12-01	15.30
1	17850.0	2010-12-01	20.34
2	17850.0	2010-12-01	22.00
3	17850.0	2010-12-01	20.34
4	17850.0	2010-12-01	20.34

A. Calculation of Total_sales

	CustomerID	frequency	recency	T	monetary_value	probability_alive	pred_num_tst
0	14911.0	131.0	372.0	373.0	1093.661679	0.999966	8.95
1	12748.0	113.0	373.0	373.0	298.360885	0.999971	7.73
2	17841.0	111.0	372.0	373.0	364.452162	0.999962	7.59
3	15311.0	89.0	373.0	373.0	677.729438	0.999964	6.10
4	14606.0	88.0	372.0	373.0	135.890114	0.999955	6.03

B. Calculation of probability_alive and pred_num_tst

	frequency	recency	T	monetary_value	probability_alive	pred_num_tst	exp_avg_sales
CustomerID							
12347.0	6.0	365.0	367.0	599.701667	0.999698	0.47	603.039974
12348.0	3.0	283.0	358.0	301.480000	0.999177	0.27	306.725566
12352.0	6.0	260.0	296.0	368.256667	0.999405	0.56	371.034740
12356.0	2.0	303.0	325.0	269.905000	0.999478	0.22	277.562082
12358.0	1.0	149.0	150.0	683.200000	0.999486	0.25	704.702786

C. Calculation of exp_avg_sales

CustomerID	frequency	recency	T	monetary_value	probability_alive	pred_num_tst	exp_avg_sales	predicted_clv	manual_predicted_clv	CLV
12347.0	6.0	365.0	367.0	599.701667	0.999698	0.47	603.039974	280.414087	283.428788	14.171439
12348.0	3.0	283.0	358.0	301.480000	0.999177	0.27	306.725566	81.591303	82.815903	4.140795
12352.0	6.0	260.0	296.0	368.256667	0.999405	0.56	371.034740	206.039238	207.779454	10.388973
12356.0	2.0	303.0	325.0	269.905000	0.999478	0.22	277.562082	59.125051	61.063658	3.053183
12358.0	1.0	149.0	150.0	683.200000	0.999486	0.25	704.702786	174.551030	176.175697	8.808785

Calculation of CLV

V. CONCLUSION

In this project work, BG or NBD and Gamma-Gamma models were used to predict Customer Lifetime Value for online retail purchasing. The BG or NBD-model, mainly focuses on the transaction of customers (alive) and predict their involvement over a period of 30 days. Here, the comparison is made between actual transactions and the predicted transactions. It has error rate of 2.8% which is comparatively less than other models. This model also predicts individual future transactions. The Gamma-Gamma model is used to predict the net future profit to the company. Here, the customers have been classified into different segments based on the net profit and duration of their involvement in the company.

REFERENCES

[1]. A. Mammadzada, E. Alasgarov and A. Mammadov, "Application of BG / NBD and Gamma-Gamma Models to Predict Customer Lifetime Value for Financial Institution," 2021 IEEE 15th International Conference on Application of Information and Communication Technologies (AICT), 2021, pp. 1-6, doi: 10.1109/AICT52784.2021.9620535.

[2]. A. P. Mauricio, J. M. M. Payawal, M. A. Dela Cueva and V. C. Quevedo, "Predicting Customer Lifetime Value through Data Mining Technique in a Direct Selling Company," 2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA), 2016, pp. 1-5, doi: 10.1109/ICIMSA.2016.7504027.

[3]. Chamberlain, Benjamin & Cardoso, Ã,ngelo & Liu, C.H. & Pagliari, Roberto & Deisenroth, Marc. (2017). Customer Lifetime Value Prediction Using Embeddings. 1753-1762. 10.1145/3097983.3098123.

[4]. H. Jia and C. Li, "The Research of Customer Lifetime Value Related to Risk Factors in the Internet Business : Exampled by Online Car-hailing Industry," 2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS), 2019, pp. 105-110, doi: 10.1109/ICIS46139.2019.8940315.

[5]. Jasek, Pavel; Vrana, Lenka; Sperkova, Lucie; Smutny, Zdenek; Kobulskx, Marek. Journal of Business Economics and Management; Vilnius Vol. 20, Iss. 3, (Apr 2019): 398-423. DOI:10.3846/jbem.2019.9597

[6]. L. C. Yean and K. T. V. Khoo, "Customer relationship management: Computer-assisted tools for customer lifetime value prediction," 2010 International Symposium on Information Technology, 2010, pp. 1180-1185, doi: 10.1109/ITSIM.2010.5561487.

[7]. L. C. Yean and V. K. T. Khoo, "Customer Relationship Management: Lifecycle of Predicting Customer Lifetime Value," 2010 Second International Conference on Computer Research and Development, 2010, pp. 88-92, doi: 10.1109/ICCRD.2010.24.

[8]. T. T. Win and K. S. Bo, "Predicting Customer Class using Customer Lifetime Value with Random Forest Algorithm," 2020 International Conference on Advanced Information Technologies (ICAIT), 2020, pp. 236-241, doi: 10.1109/ICAIT51105.2020.9261792.

[9]. X. N. Wang, "Research of CLV Forecasting Model Based on E-Commerce Website," 2009 First International Conference on Information Science and Engineering, 2009, pp. 3002-3005, doi: 10.1109/ICISE.2009.906.

[10]. Yi Wang, S. Sanguansintukul and C. Lursinsap, "The customer lifetime value prediction in mobile telecommunications," 2008 4th IEEE International Conference on Management of Innovation and Technology, 2008, pp. 565-569, doi: 10.1109/ICMIT.2008.4654427.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 6, June 2022

BIOGRAPHY

[1]. Jake Vander plas, "Python Data Science Handbook: Essential tools for working with data", O'Reilly Publishers, I Edition

[2]. Allen B. Downey, "Think Python: How to Think Like a Computer Scientist'', 2nd edition, Updated for Python 3, Shroff/O'Reilly Publishers, 2016