Predicting Customer Lifetime Value using ML: A Comparative Analysis

Anand Kumar N
Reva Academy for Corporate Excellence – RACE, REVA University,
Bangalore, Karnataka 560064
anandn.ba06@reva.edu.in

Mithun DJ
Reva Academy for Corporate Excellence - RACE, REVA University,
Bangalore, Karnataka 560064
mithun.dj@reva.edu.in

Rashmi Agarwal
Reva Academy for Corporate Excellence – RACE, REVA University,
Bangalore, Karnataka 560064
rashmi.agarwal@reva.edu.in

Abstract *- For the businesses to at-least sustain in competitive markets, the key is to identify such categories of clients and target only the clients that add more value. Brand awareness, client attitudes, or even sales and shares that fall under the marketing metrics category are not good enough to show a return on the investment in marketing the offering as these are traditional in nature. This drives the need to predict the Customer Lifetime Value (CLTV) for a business. Financial metrics like the aggregate profit of a business or stock price which are measured, are useful with diagnostic capability. Recent studies have come up with the observations that there will be no uniformity across all clients being profitable. It may be a wise strategy to reach out to target some clients or allocate resources differently to specific segments of clients. CLTV is a key metric that can be predicted for a customer to know the future value that the customer can add to a business or firm. The value is realized in terms of the profit a firm can make from all the purchases a customer makes with the firm. Additionally, it signifies the long-term relationship the customer has with a firm. As part of this study, CLTV is predicted using various linear and non-linear machine learning techniques. Predicted CLTV is being compared based on the key KPIs to identify the best technique that could be used. Based on the CLTVs, customers can be categorized into different groups like most profitable to least profitable. This categorization helps the marketing teams to come up with strategies that enable the business to tap the full profit potential of the customers. Success of a firm depends on the firm's ability to attract and retain customers that are loyal and valued to the business.*

Keywords – CLTV, CART, CHAID, ExCHAID, MLP, RBF, KNN, QUEST

I. INTRODUCTION

Customer Value or Customer Lifetime Value (CLTV) is picking up a lot of attention as a metric in the marketing domain of business. CLTV has been used by businesses like Retail, Telco and others as a measure of success for a business. Business have pressure to make marketing accountable. Businesses are not able to realize the Return on Investment (ROI) (Marketing Investment) by using traditional metrics like brand awareness, attitudes, shares and stocks. Marketing actions to improve sales or shares could have an impact on the long term profitability of the business [1].

Financial metrics like aggregate profit of a business or stock price which are measured are useful with diagnostic capability. Recent studies in the area customer value measurement have come up with the observations that there will be no uniformity across all clients being profitable. It may be a wise strategy to reach out to target some clients or allocate resources differently to specific segments of clients.

Resources can be allocated by identifying customers who are profitable based on prediction of CLTV, a metric that is disaggregate in nature. Technology advancements are enabling the businesses gather infinite amount of client activity data which lets them use this data as preferred rather than intentional. Modelling sophistications have enable marketers to convert the data into valuable insights.

Nowadays, 69% of firms monitor CLTV, but they do it inefficiently. Instead, 81% of firms are doing an extraordinary job in gauging customer value to expand their selling power. As per a study conducted by one of the firms, 55% of companies that are developing think that it is "Very important" to invest in customer focused service programs. Observations from another company showed that a getting the retention rate up by 5% could result in a 25% to 95% expansion in profits [2].

Equation (1) mentioned below is used to calculate CLTV [3].

$$CLTV = \sum_{t=0}^{T} \frac{(p_t - c_t)\, r_t}{(1+i)^t} - AC \qquad (1)$$

Where

p_t = price paid by a consumer at time t,

c_t = direct cost of servicing the customer at time t,

i = discount rate or cost of capital for the firm,

r_t = probability of customer repeat buying or being "alive" at time t,

AC = Acquisition cost, and

T = time horizon for estimating CLTV.

The Equation 1 used to calculate CLTV looks simple but the complexity around it will unfold as exploration of the definition of CLTV picks up. Future sales and values in monetary terms is being predicted by making use of the client purchase data that is historic in nature brings in the complexity to some extent.

As part of this study, calculating the CLTV using Linear and Non-Linear modelling methods is being carried out.

Then comparing the R squared values from each of these methods to conclude as which method is a best fit for predicting the CLTV for the retail apparel business.

II. LITERATURE REVIEW

CLTV means the value that a client adds to the business which gauges all the potential profits a particular client can bring to the organization. For example, let's consider an online shop selling sports goods and all the additional products, and a new client has just purchased a bat. In the future, they may purchase a ball, wicket, gloves, etc. At some point, they may come for another bat. All these potential purchases and revenues are the value that customer would add.

First, the prediction of CLTV is carried out at an individual client or segment level. This lets the businesses differentiate between clients who are of more value than others rather than simply examining an average across profit. Second, unlike finance, CLTV explicitly incorporates the possibility that a customer may defect to competitors in the future [3].

This value is one of the most important factors and plays a vital role when it comes to maximizing the company's efficiency. The benefits of CLTV have been already mentioned. However, here is a more detailed example: when the total cash flow of a given customer is known, it is straight forward to understand how far the business has got with customer retention and maximize Return On Investment (ROI) [2].

Value of customer through the lifetime with a business is something that can be gauged in terms of monetary is the total value that a client adds to the business from all the purchases they make his entire time of relationship. The lifetime in this context is the entire time the client is purchasing from the business before moving to competitors [4].

There are numerous CLTV models that have been developed. These models are the PCV model (Past Customer Value), RFM model (Recency, Frequency, Monetary), SOW model (Share of Wallet) and future-past customer behaviour model [5].

Future behaviour of customers is considered by the future - past customer behaviour models whereas some analytical models include acquisition cost when calculating lifetime value while some others do not. Retention rate is used by most of the future-past customer behaviour studies to determine the activation period.

Different methods like generalized regression, logistic regression, quantile regression, latent class regression, CART, Markov chain modelling, neural network to create past customer behaviour models, etc. are used by many studies [6].

Techniques like decision trees, clustering, logistic regression, artificial neural networks, support vector machine, random forests, etc. are used by industries like retail, insurance, banking, telecommunication, financial services taking advantage by significantly using data mining techniques for identifying CLTV and performing analysis based on CLTV. Various linear and non-linear modelling techniques are used to model the data to predict CLTV and then get the R-squared value along with mean error. Linear algorithm like Logistic Regression, Special Variable Selection methods and non-linear algorithms like the decision trees – Chi Square Automatic Interaction Detection (CHAID), Classification and Regression Trees (CART), Quick Unbiased Efficient Statistical Tree (QUEST), neural networks like the Multi Layered Perceptron (MLP), Radial Basis Function (RBF) along with KNN are being used to derive the value. Adjusted R-squared value along with Mean Absolute Error are compared to determine the best fit model that can be used to predict the CLTV.

Other methodologies that are available like the proposed extended RFM analysis method with one additional parameter called Count Item can be used. When the results of these approaches are compared, it's understood that there is no difference to the clustering result with addition of the count item as a new parameter to the RFM method. So, the weighted RFM method is used to calculate CLTV for each segment. Marketing and sales strategies of the company can be explained by the results of calculated CLV for different segments. Probabilistic models like Pareto/NBD (Negative Binomial Distribution) and Beta Gamma/Negative Binomial Distribution (BG/NBD) are powerful techniques for predicting the future activity of a customer.

There are numerous machine learning modelling techniques that can be used to predict CLTV, in this paper around 10 machine learning techniques have been used to predict CLTV and their Adjusted R-squared values and Mean Absolute Error have been compared to identify the best technique that can be made use of.

III. OBJECTIVES OF THIS STUDY

Below are the objectives for predicting CLTV:

1. Compute CLTV for the customers using Linear and Non-Linear machine learning algorithms
2. Compare the Adjusted R-squared values derived through regression using the CLTVs to identify the best modelling technique
3. Deploy the best modelling technique for prediction of CLTV based on the comparison outcomes

Firms can use insights derived from the data to segment their customers, this helps in customer segmentation. Segmentation will enable the firms to assess the customer's loyalty along with the projected revenue. Strategies that are effective and efficient needs to be designed to keep the profitable customers intact.

IV. METHODOLOGY

The methodology followed is CRISP-DM framework that starts with understanding the business and then narrowing down into specific areas of interest like understanding the data, preparing the data for more insights, building the model, evaluating the model and deployment.

Looking into various studies, it's understood that there are two broad approaches to modeling the CLTV problem [4]

1. Historical Approach:
   i. *Aggregate Model*: this method calculates the CLTV based on past purchases taking into consideration the average of revenue per customer, it gives us a single value for the customer.
   ii. *Segmentation Model*: this method groups the customers into different segments based on the transaction date, etc., and calculates the average revenue per segment. This method gives CLTV value for each segment.
2. Predictive Approach:
   i. *Model based on ML*: this makes use of different regression techniques to fit on past data to predict the CLTV.
   ii. *Probabilistic Model*: estimates the count of purchases to happen in the future with monetary value for each purchase by making use of the data distribution based on probability

As part of this paper, CLTV is predicted using Linear Modeling Techniques like Logistic Regression along with Variable Selection methods under Logistic Regression and Non-Linear modelling techniques like Decision Tree Methods – CART, CHAID, Exhaustive CHAID, and Quest. Neural Networks – MLP and RBF and KNN. Then the R-squared is derived from the predicted CLTV values using regression. The Adjusted R-squared values from each of these techniques are compared to identify the best machine learning technique to predict the CLTV.

The main target variable that needs to be predicted here by looking at the data is if a customer repeats his purchase or not with the retailer. If the customer repeats his purchases, then the CLTV value would be on the higher side. As part of predicting the CLTV, it is more interesting to identify the repeat customers who give more business.

In-order to identify the customers that are of high value to the business, there is a need to identify the customers that churn meaning who do not repeat their purchases after their initial or first purchase. Using Logistic Regression machine learning algorithm, derive the Churn Propensity which is used to arrive at the Churn Indicator based on the first purchase date and the recent purchase date. Using this Churn Indicator, come to a conclusion if a customer has Churned or not. Churn is an important flag type variable whose value helps group the customers into repeat or non-repeat customers leading the way to calculate the CLTV for those repeat customers. CLTV is calculated on a monthly basis using the formula discussed in the Data Understanding section for the customers. Here multiple machine learning techniques have been used to see which model can fit the best with the given data.

A. BUSINESS UNDERSTANDING

Focus for this paper is on retail business. Retail business is more challenging than ever. Competition is exponential along with Amazon being a major player in this domain which drives the acquisition costs to shoot up while customer bringing down the profitability. It's a double edged sword. The days of increasing the size of the businesses by acquiring new customers are gone, it is the value that these customers can bring is what matters. To increase revenue from the existing client base and to get a deeper understanding of the clients with high value is the only way today for the retailers to grow the business. Keeping the customers repeat their purchases is the key to get more value. "In the moment" personalization, using technology like black box which is gaining a lot of momentum to convey the right information at the correct point of time to each client. Getting to know for sure who the customer is and what they intend to do by matching the customer journeys with insights is the key driving factor of customer engagement that brings in value.

B. DATA UNDERSTANDING AND PREPARATION

Data is the most important component that will help understand the customer behavior and gain insights. Data needs to be thoroughly examined and understood to derive meaningful outcomes. Data used in this study is from a retail apparels business that sells apparels across India. This is set contains all the purchases occurring between 2018 and 2020 for a period of 3 years.

Attribute Information - Some of the key attributes are listed here to get a glimpse into the dataset: Order No., Order Date, Customer ID, Quantity, Unit Price, Status.

There is a need to see if there is any need for cleanup of the data by looking for duplicate records, null value rows, etc. Checking for missing values in the data as part of preparing for data modeling. For the Lifetime value calculation, make use of the existing columns and calculate the parameters as per the formula to get the target variable - CLTV.

Listed here are the key features that are created to calculate:

1. t - this is value derived from the difference of recent order date and first order date
2. Sum of Product Discounted – This is based on the discount rate
3. Acquisition Cost = 0 (0 has been assumed as this is an online retailer)
4. Churn Propensity which is used to calculate Alive Propensity
5. Alive Propensity is 1 – Churn Propensity
6. Order Month – This is calculated based on the month of the order date

C. MODELLING

As part of this study, an extensive analysis has been carried out by predicting the CLTV using various linear machine learning techniques like Logistic Regression and Variable Selection methods under Logistic Regression like Forward Likelihood Ratio, Backward Likelihood Ratio, Forward Condition and Backward Condition. Non-linear machine learning techniques namely: Decision Tree – CHAID, CART, Quest () and ExChaid (Exhaustive CHAID). Neural Networks – MLP and RBF and KNN.

1. Logistic Regression:

Logistic regression is a supervised learning algorithm used when the target (dependent) variable is categorical. It is used to predict a binary (yes/no) event occurring. In this case, logistic regression is used to predict if the customer will churn or not. Churn indicator is an important metric that is being used here to classify customers based on their purchase. Churn propensity estimates the likelihood of a customer to leave in the next period of time. In our case, churn propensity is based on if the customer has repeated purchasing from the retailer or not. If the customer has purchased only once and has not purchased anything again, then the customer is considered as churned and if the customer has repeat purchases, then the customer is considered as not churned. For calculation of CLTV using the equation, alive propensity is needed, as the Churn propensity has been derived as part of the logistic regression model. Alive propensity can be calculated by using 1-Churn propensity. Once the customers are classified into two groups as Churned and Not Churned, predicting the CLTV for all those customers that are classified as Not Churned.

2. Variable Selection

This is a process in which it starts with all the variables being considered part of the equation and are then removed one by one. The one having the smallest partial correlation with the dependent variable is considered for removal first. If it satisfies the condition for removal then it is removed [8].

- Backward Elimination

Backward Elimination is a simple method of all the variable selection methods. This method considers all the variables in the model and start with a full model. The variables are removed one by one from the model until the remaining variables are considered to have significant contribution on the outcome. In Backward Elimination, there are Backward Conditional and Backward Likelihood Ratio methods.

- Forward Selection

Forward selection is the opposite of the backward elimination method. It starts with an empty model without any variables and then the addition of variables is carried out until any variable not added to the model can have significant impact on the outcome of the model. In Forward selection, there are Forward Conditional and Forward Likelihood Ratio methods.

The variable selection methods that are being used here to model the predicted CLTV have different adjusted R-squared values. Forward Conditional and Forward Likelihood techniques have a higher R-squared value compared to the Backward Conditional and Backward Likelihood. It can be inferred that as the forward techniques are performing better due to the fact that the model starts with no variables in the model and then adds one by one until any variable not included in the model can have a significant contribution to the model whereas the backward method starts with a full model and proceeds with removal of variables one by one until all remaining variable have significant contribution to the model outcome.

3. Decision Trees

Decision trees are one of the supervised learning techniques which can be used for both classification and regression problems, preference is for classification problems. It is a tree-structured classifier, features of a dataset are represented by internal nodes, branches represent the decision rules and each leaf represents an outcome.

- CHAID

CHAID is an abbreviation for Chi Square Automatic Interaction Detection modeling technique that is one of the decision tree methods. It is a statistical technique primarily used in market research.

- CART

CART is the abbreviation for Classification and Regression Trees that is one of the decision tree modeling techniques. CART is a predictive algorithm which predicts the value of a target variable based on other variables.

- Exhaustive CHAID

ExCHAID is a modification to the basic CHAID algorithm that performs a more thorough merging and testing of predictor variables.

- QUEST

Quest stands for Quick, Unbiased and Efficient Statistical Tree. Quest method provides selection of features unbiased and categorical variables are handled with several categories.

Looking at each of the decision tree modeling techniques, ExCHAID with an error of 284 is the best value and is a good fit to the data when compared to QUEST, CHAID and CART. It can be observed that each of the decision tree techniques model the data with a different approach which leads to the difference in the outcome.

4. Neural Networks

Neural Networks are a series of algorithms that mimics the way human brain works in identifying the relationships in a dataset [9].

- Multi Layered Perceptron

In MLP, perceptron's are arranged in interconnected layers. The input layer collects input patterns. The output layer has classifications or output signals to which the input patterns may map. Hidden layer fine tune the input weightings until the neural networks margin of error is minimal.

- Radial Basis Function (RBF)

RBF is a type of neural network that is commonly used. A RBF network is forward feed type of neural network that consists of three layers, input layer, hidden layer and the output layer.

Looking at the performance of the two neural network models being used here – Multi Layer Perceptron and Radial Basis Function, Radial Basis Function with a lesser error value of 336.35 looks to be a better model compared to Multi-Layer Perceptron with a value of 358.59.

5. K-nearest neighbors (KNN)

KNN is a non-parametric, supervised learning, classification algorithm which uses proximity to make classifications or predictions.

KNN model has a lower Mean Absolute Error value compared to other techniques and hence can be considered as the one that fits well to the data. An extensive analysis has been performed using all the above modelling techniques to predict the CLTV.

Adjusted R-squared values generally ranges from 0 to 1. Adjusted R-squared has been used as a key metric for each of the linear techniques as this value determines the variance in the dependent variable that can be explained by the independent variable. It shows how well the data fit the model or the goodness of the fit. Mean Absolute Error is used a metric in case of non-linear techniques. A lower MAE indicates a good fit model. Although the Adjusted R-squared provides some useful insights into the model, it does not disclose the causation relationship between the independent and dependent variables. There is no universal rule on how to incorporate the statistical measure in assessing a model [10].

V. FINDINGS

CLTV has been calculated or predicted using Linear and Non-Linear machine learning techniques. CLTV thus predicted are used to derive the Adjusted R-squared value using which can be compared to different models to find the best model that fits well to the data. Then this model can be used by the business to predict CLTVs which enables the marketing decisions to target the high value customers who could increase the profit margins of the firm.

In case of linear techniques, "Fig. 1" shows the comparison between them.

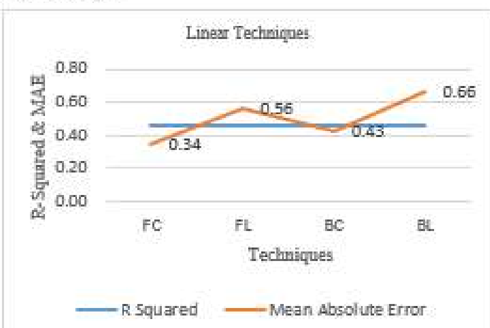

Fig. 1. Linear techniques comparison

"Fig. 2" shows the comparison of non-linear techniques

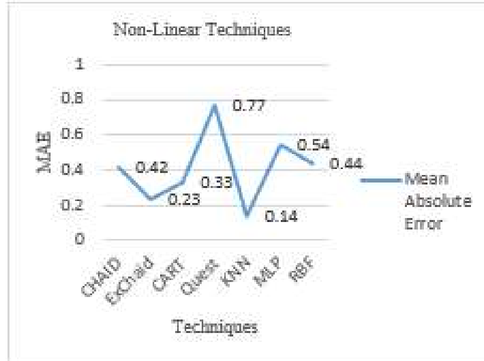

Fig. 2. Non-linear techniques comparison

Table I shows the comparison of all the techniques discussed as part of this study with their key metrics.

TABLE I. Comparison of Techniques

Algorithm	Adjusted R-Squared	MAE	MAPE	Min-APE	Max. APE	RMSE
Forward Conditional	0.46	0.34	19%	15%	25%	0.9
Forward Elimination	0.46	0.56	27%	18%	30%	0.6
Backward Conditional	0.45	0.43	30%	20%	35%	0.01
Backward Elimination	0.45	0.66	45%	25%	50%	0.6
CHAID	NA	0.42	65%	33%	70%	0.3
ExCHAID	NA	0.23	48%	30%	60%	0.2
CART	NA	0.33	36%	25%	45%	0.8
QUEST	NA	0.77	38%	20%	50%	0.5
KNN	NA	0.14	8%	5%	13%	0.4
MLP	NA	0.54	33%	20%	40%	0.03
RBF	NA	0.44	32%	25%	45%	0.01

From the above comparison charts, it can be observed that KNN has less error compared to other techniques. Hence KNN can be selected as the best fit model. Based on this the retail firm can allocate budget for retaining the high value customers by offering good discounts.

VI. CONCLUSION

The marketing team can now make use of the CLTV values to target high value customers or group of customers to drive more sales and profit.

Also, it is hard for the firms to reach individual customers. This calls for the need to segment customers based on demographic data if it was available. Segmentation based on demographics could provide more insights into the customer profile to focus more on the customers. The CLTV calculated or predicted using different modeling techniques helps the firm to take decision in terms of promotions and other offers that can be extended to their high value customers.

Customer value can be used to come up with business plans that can be effective in terms of driving the value add from each of the customers along with providing room for scaling the business. The strategy defined by the marketing teams can indicate the profit extracted. In general, customer lifecycle is managed by the automation platform under marketing. Marketing campaigns are orchestrated by the marketing platforms which also automate the movement of leads and the client pipeline. The planning, coordination, execution, management and measurement of the campaigns and automation of various repetitive tasks are carried out by using these software applications.

The CLTV calculated or predicted helps the firm to take decision in terms of promotions and other offers that can be extended to their high value customers. The objective of predicting CLTV for retail firms is to ensure that the firms are in a state of mind to know who their high value customers are and can accordingly work on retaining them to drive higher profit margins.

The CLTV should start the shift from product-centric to customer-centric approach. With the increase in awareness of the CLTV metric and its benefits, the focus will be on its widespread adoption. Implementation of CLTV based strategies will give good insight to improve the customer experience

Some recommendations to increase customer Life Time Value are -

1. Effective Communication
2. Loyalty Program
3. Retargeting

VII. REFERENCES

[1] L. Singh, N. Kaur, and G. Chetty, "Customer Life Time Value Model Framework Using Gradient Boost Trees with RANSAC Response Regularization," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2018-July, 2018, doi: 10.1109/IJCNN.2018.8489710.

[2] "Customer Lifetime Value Prediction using Machine Learning | Addepto," 2019. https://addepto.com/customer-lifetime-value-prediction-machine-learning/ (accessed Aug. 26, 2022).

[3] B. Hardie, "Modeling Customer Lifetime Value," *J. Serv. Res.*, vol. Volume 9, no. Clv, p. 17, 2006, doi: 10.1177/1094670506293810.

[4] "Predicting Customer Lifetime Value : A Definitive Guide," 2020, Accessed: Mar. 23, 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/10/a-definitive-guide-for-predicting-customer-lifetime-value-clv/

[5] S. Sharma, "Customer Lifetime Value Modelling," pp. 1–31, 2021.

[6] "How Retail Brands Can Predict Customer Lifetime Value," 2019. https://tinuiti.com/blog/ecommerce/customer-lifetime-value/ (accessed Mar. 23, 2022).

[7] A. Manasson, "Why using CRISP-DM will make you a better Data Scientist | by Alexander Manasson | Towards Data Science," 2019. https://towardsdatascience.com/why-using-crisp-dm-will-make-you-a-better-data-scientist-66efe5b72686 (accessed Aug. 26, 2022).

[8] M. Z. I. Chowdhury and T. C. Turin, "Variable selection strategies and its importance in clinical prediction modelling," *Family Medicine and Community Health*, vol. 8, no. 1. 2020. doi: 10.1136/fmch-2019-000262.

[9] X. Pang, Y. Zhou, P. Wang, W. Lin, and V. Chang, "An innovative neural network approach for stock market prediction," *J. Supercomput.*, vol. 76, no. 3, pp. 2098–2118, Mar. 2020, doi: 10.1007/S11227-017-2228-Y.

[10] CFI, "R-Squared - Definition, Interpretation, and How to Calculate," 2019. https://corporatefinanceinstitute.com/resources/knowledge/other/r-squared/ (accessed Aug. 27, 2022).