REVA UNIVERSITY
Bengaluru, India

A Project Report on

Customer Life Time Value with Machine Learning – A Comparative Study

Submitted in Partial Fulfilment for Award of Degree of

Master of Business Administration
In Business Analytics

Submitted By

Anand Kumar N

R19MBA52

Under the Guidance of

Mithun DJ

Senior Manager – Data Science,
RACE, REVA University

REVA Academy for Corporate Excellence - RACE

REVA University

Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064

race.reva.edu.in

August, 2022

Candidate's Declaration

I, **Anand Kumar N** hereby declare that I have completed the project work towards the second year of Master of Business Administration in Business Analytics at, REVA University on the topic entitled **Customer Life Time Value with Machine Learning – A Comparative Study** under the supervision of **Mithun DJ**. This report embodies the original work done by me in partial fulfilment of the requirements for the award of degree for the academic year **2022.**

Place: Bengaluru

Name of the Student: Anand Kumar N

Date: 27 Aug 2022

Signature of Student

Certificate

This is to Certify that the project work entitled **Customer Life Time Value with Machine Learning – A Comparative Study** carried out by **Anand Kumar N** with **R19MBA52,** is a bonafide student at REVA University, is submitting the second-year project report in fulfilment for the award of **Master of Business Administration** in Business Analytics during the academic year 2022. The Project report has been tested for plagiarism and has passed the plagiarism test with the similarity score less than 15%. The project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said degree.

Signature of the Guide

Name of the Guide: **Mr. Mithun DJ**

Signature of the Director

Name of the Director: **Dr. Shinu Abhi**

External Viva

Names of the Examiners

1. Dr. Sai Hareesh, Research Expert, SAP Labs India
2. Pradeepta Mishra, Director - AI, L&T InfoTech

Place: Bengaluru

Date: 27 Aug 2022

REVA UNIVERSITY
Bengaluru, India

Acknowledgement

I am highly indebted to **Dr. Shinu Abhi**, Director, and Corporate Training for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project.

I would like to thank my project guide **Mr. Mithun D. J.** for the valuable guidance provided to understand the concept and execute this project. It is my gratitude towards **Dr. Jay Bharateesh Simha** and all other mentors for the valuable guidance and suggestion in learning various data science aspects and for their support. I am thankful to my classmates for their aspiring guidance, invaluable constructive criticism, and friendly advice during the project work.

"I would like to acknowledge the support provided by the founder and Hon'ble Chancellor, **Dr. P Shayma Raju**, Hon'ble Vice-Chancellor, **Dr. M. Dhanamjaya**, and Registrar, **Dr. N. Ramesh**."

It is my sincere gratitude towards my parents, and my family for their kind co-operation and encouragement which helped me in the completion of this project.

Place: Bengaluru

Date: 27 Aug 2022

REVA UNIVERSITY
Bengaluru, India

Similarity Index Report

This is to certify that this project report titled **Customer Life Time Value with Machine Learning – A Comparative Study** was scanned for similarity detection.Process and outcome are given below.

Software Used: Turnitin

Date of Report Generation: 27 Aug 2022

Similarity Index in %: 13%

Total word count: 7368

Name of the Guide: Mithun DJ

Place: Bengaluru

Name of the Student: Anand Kumar Narayanappa

Date: 27 Aug 2022

Signature of Student

Verified by: Dincy Dechamma

Signature

Dr. Shinu Abhi,

Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	CLTV	Customer Lifetime Value
2	AC	Acquisition Cost

List of Figures

List of Tables

List of Equations

Abstract

For the businesses to at-least sustain in competitive markets, the key is to identify such categories of clients and targeting 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 for predicting the Customer Lifetime Value (CLTV) for a business. In the process of making internal decisions like: from allocating budget marketing spends, to minimizing potential losses and preventing customer churning, CLTV plays a key role. The value that a client adds through his lifetime with a business represents the total amount of money a customer spends in a given business or product during their life time which means time during which the customer is actively purchasing from that business. In a big picture, the customer value can be looked up as a measure of the profit associated with a particular customer relationship, which should be the guiding factor on how much the business is interested in retaining that customer or customer segment.

In order to foster the full potential of loyal customers, the customers need to be first identified based on their value to the business and then segmented to group them into targeted and profitable customers, prediction of CLTV is the initial point. The ability of an organization to identify, create, nurture and sustain a loyal and high valued client experience and relationships is an important aspect of corporate success.

The proposed solution framework here provides a step-by-step process of calculating CLTV for a retail apparel business using different methods of prediction. CLTV is a recent marketing paradigm that helps to pursue long term relationships with profitable customers.

As part of the solution, CLTV is predicted with the equation by making use of various Linear and Non-Linear machine learning algorithms.

Linear machine learning algorithms like Logistic Regression, Special Variable Selection methods under Logistic Regression like the Back Likelihood Ratio, Forward Likelihood Ratio, and Backward Conditional and Forward Conditional techniques have been made use of.

Non-linear machine learning techniques namely: Decision Tree – CHAID (Chi Square Automatic Interaction Detection), CART (Classification and Regression Trees), Quest (Quick Unbiased Efficient Statistical Tree) and ExChaid (Exhaustive CHAID). Neural Networks – MLP (Multi-Layer Perceptron) and RBF (Radial Basis Function) and KNN – K Nearest Neighbors are being used to predict the CLTV and then derive the R-Squared value to determine the best model fit to the data.

Solution and findings for this project.

Keywords: CLTV, CLV, Customer Value, Sentiment Analysis, MLP, RBF, CHAID, ExCHAID, CART, KNN,

Contents

Chapter 1: Introduction

This chapter provides a brief introduction to the concept of CLTV – Customer Lifetime Value and its importance in increasing the profit margin of businesses. It also introduces the formula that is being used to predict the CLTV for the retail apparel business.

Customer Value or Customer Lifetime Value 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

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.

With making use of the current technological enhancements, it enables the businesses to use these insights to introduce custom marketing programs for individual clients or segments of clients.

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 (*Customer Lifetime Value Prediction Using Machine Learning | Addepto*, 2019) (*Customer Lifetime Value Prediction Using Machine Learning | Addepto*, 2019).

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.

Calculating CLTV for different customers helps in several ways, mainly regarding business decision making. Knowing CLTV, many things can be determined but not limited to some of the below key items(Sharapa, 2019):

a. To calculate how much a business needs to spend to acquire a similar customer and have a relationship that is profitable

b. To predict the kind of products customers with the highest CLTV want?

c. Identify the highest profitability products

d. Identify the most profitable clients

e. Together, these decisions can significantly boost profitability

Calculating or Predicting CLTV is an important step towards increasing the profitability of the business for a firm:

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.

The general equation to calculate CLTV is as below: (Hardie, 2006).

$$CLTV = \sum_{t=0}^{T} \frac{(p_t - c_t)\, r_t}{(1+i)^t} - AC \qquad (1.1)$$

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 1.1 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 project, 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.

This chapter provided an introduction along with the equation to predict CLTV. The next chapter is all about the literature review of the CLTV concept by reviewing the various research papers that are being referred as part of this exercise.

Chapter 2: Literature Review

The second chapter here talks about the literature review of the CLTV concept by referring the various research papers not just limited to the ones listed in the bibliography.

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.

CLTV is generally defined as the present value of all future profits obtained from a customer over his or her life of relationship with a firm. CLV is like the discounted cash flow approach used in finance. However, there are two key differences.

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 (Hardie, 2006).

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 ROI (return on investment) (*Customer Lifetime Value Prediction Using Machine Learning | Addepto*, 2019).

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

(Predicting Customer Lifetime Value: A Definitive Guide n.d.) (*Predicting Customer Lifetime Value : A Definitive Guide*, 2020).

A company can come up with strategies by which they can retain their clients and the ability to increase the overall profit by understanding the customers. Companies can investigate the parameters that companies generally ignore by predicting CLTV. At the beginning of a relationship, customers are more valuable due to the future potentials that they offer.

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 (Sharma 2021).

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 (*How Retail Brands Can Predict Customer Lifetime Value*, 2019).

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 – CHAID, CART, QUEST, neural networks like the MLP, RBF along with KNN are being used to derive the value.

The R-squared value along with errors 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 and 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 project around 10 machine learning techniques have been used to predict CLTV and their R-squared values have been compared to identify the best technique that can be made use of.

Once the literature review of the CLTV concept is completed, the next important step is to define the problem. The next chapter states the problem that will be addressed by predicting the CLTV.

Chapter 3: Problem Statement

The key problem areas that are addressed as part of this project work is

1. Determine the best model fit to predict CLTV by exploring various linear and non-linear machine learning algorithms

2. Perform a comparison of the key metrics of these techniques to determine the best fit model to recommend for prediction of CLTV

3. Calculate the monthly average value of customers for business to drive their marketing strategies

Chapter 4: Objectives of the Study

As part of this chapter, the key objectives to be achieved are listed down.

Below are the objectives for predicting CLTV:

1. Compute CLTV for the customers using Linear and Non-Linear machine learning algorithms

2. Compare the R-squared values derived through regression using the CLTVs to identify the best modelling technique

3. Deploy the 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.

Listing the key objectives is an important activity, with that completed. The next chapter provides details about the project methodology.

Chapter 5: Project Methodology

Project methodology that will be used as part of this project is being discussed in this chapter with details of the approach.

The process that will be followed as a project methodology 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. Figure No. 5.1 is the pictorial representation of CRISP-DM phases with the flow (Manasson, 2019).

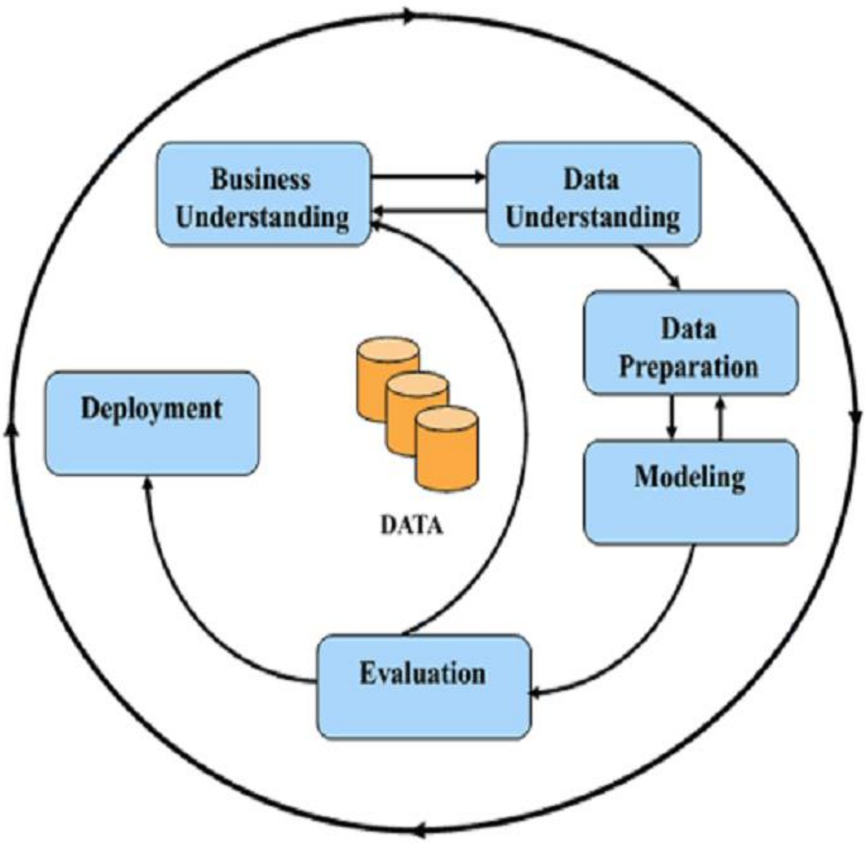

Figure No. 5.1 CRISP DM Methodology(Manasson, 2019)

Looking into various studies, its understood that there are two broad approaches to modeling the CLTV problem(*Predicting Customer Lifetime Value : A Definitive Guide*, 2020).

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 method 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 project, 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 – Multi Layer Perceptron and Radial Basis Function and K Nearest Neighbors. Then the R-squared is derived from the predicted CLTV values using regression. The R-squared values from each of these techniques are compared to identify the best machine learning technique to predict the CLTV.

Below are the details about the approach that has been used to predict CLTV for the Apparel_POS data: Data used for this project consists of all the purchases from 2018 to 2020 for a period of 3 years. The company is an Apparels Retailer that sells apparels across India.

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

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.

As the project methodology and approach has been covered in this chapter, the next chapter onwards, each of the steps part of the methodology are covered in more detail. To start with, the next chapter elaborates the Business Understanding area.

Chapter 6: Business Understanding

Here the first step of the CRISP DM methodology, Business Understanding has been elaborated along with this relevance to the prediction of CLTV.

As part of this project, focus 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.

The websites and algorithms part of the recommendation engines for products available today are not in position to deliver this kind of engagement with the customers. To understand the client's preferences and behaviors, the next level of analytics that is advanced in functionality and machine learning techniques needs to be made use of.

Activation of these insights in an easier way that enables retailers to attract more clients that are of higher value, increase loyalty within the existing client base and step in at the right time when clients deviate from their expected purchase journey is what is needed from a system that would be used.

To achieve the business goals like transforming the buyers who purchase for one time to purchase again and again, decrease the number of clients that are churning out, and getting increased client value can be driven by making use of predictive analytics that combines details of clients, steps and quantification for the marketers.

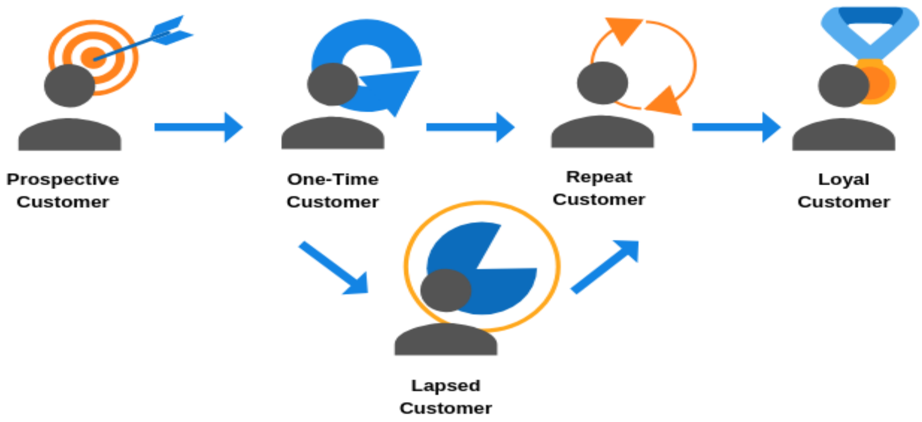

Figure No. 6.1 Customer Conversion(Burkard, 2020)

Figure No. 6.1 depicts the process of customer acquisition, retention and churn. Metrics like count of clicks on the item, scores that indicate client engagement , percentage of clients that are churning, number of items added to the cart and so on that are tracked by the marketers today are kind of distractions. The key insights like the base-line impact to the business – number of purchases and profits from those purchases can be captured making use of the metrics captured.

Retailers need to shift the thoughts to thinking about their customers holistically in addition to looking at the channels that are isolated from others and additionally utilize metrics that are channel-specific like the metrics discussed earlier to keep an eye on the performance.

"Let me go and click on that email I saw on my mobile device while I am in the store to ensure the retailer can attribute my purchase" may not be something a customer would say. Looking at the long term profitable growth by focusing on customer-centric metrics will let the retailers to come up with the best decisions.

Predicting the value add by a client is key activity to address these objectives. Knowing client value is the best and candid way to run the business.

By predicting CLTV, below are some of the key business problems that could be tried to provide an answer to:

1. Identifying the most profitable customer/s
2. How best offers could be offered by the company to make the most?
3. Drive segmentation of profitable customers
4. Budget allocation for acquiring new customers to the business
5. Retention of the existing customers by various offers and promotion

Client information that is important in nature can be pulled out, that supports making of decisions that are critical for the business by leveraging the data mining techniques.

Business can plan on changing their marketing strategy based on the CLTV predicted which in turn helps increase the profitability and their position in the market among their competitors.

Once the Business Understanding has been clearly documented, the next important step in the process is understanding the data.

Chapter 7: Data Understanding

As part of this chapter, the data is looked at very closely from different perspectives to ensure the data is all set and is in good shape for analysis and prediction of CLTV.

Data is the most important component that will help understand the customer behaviour and gain insights. Data needs to be thoroughly examined and understood to derive meaningful outcomes.

Data used in this project is from 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.

Below table provides information about the different columns that are part of the data that is being used for data mining and analysis to calculate the CLTV for the customers that have purchased apparels from this firm.

Here is the information on some of the key attributes:

i. ***Order No:*** Order Number. Nominal, an integer number uniquely assigned to each purchase. The code starts with letter 'M'.
ii. ***Quantity:*** The number of each product (item) per purchase. Numeric.
iii. ***Order Date:*** Order Date and time. Numeric, the day and time when each item was ordered.
iv. ***MRP:*** MRP. Numeric, price of each item in Rupees.
v. ***Customer_ID:*** Customer identification number. Nominal, an integer number uniquely assigned to each customer.

Data that is considered for this project has around 37453 transactions that include purchases made by different customers from across the country. This data will be prepared as part of data mining in terms of cleanup, getting rid of duplicate records if any, removal of null value rows, etc. to make it meaningful for the purpose of our analysis.

Column	Data Types
Order No	Object
External Order No	Int64
Order Date	DateTime64[ns]
Order Type	Object
Status	Object
Customer Name	Object
Country	Object
State	Object
City	Object
Email	Object
Color	Object
Quantity	float64
Return Qty	float64
Order Currency	Object
Price	Object
Ship Cost	Float64
Packing Cost	Float64
Discount	Object
Discount Code	Object
Tax	Float64
Invoiced	Object
Base Currency	Object
COGS	Object
Invoiced In Base Currency	Object
Gross Margin	Object
GM Percent	Float64
On Hold Status	Object
Replacement Order	Object
Primary color	Object
MRP int64	Object
Product Discounted?	Object
Product Discounted %	Object
Account object	Object
New order No.	Object
Customer_ID	Int64
State_City	Object
Final state	Object
Final City	Object

Table No. 7.1 Attributes and Data types

As per the Equation 7.1 for calculating CLTV : (Hardie, 2006)

$$CLTV = \sum_{t=0}^{T} \frac{(p_t - c_t)\, r_t}{(1 + i)^t} - AC \qquad (7.1)$$

Where

p_t = price of the item that a consumer pays at time t,

c_t = direct cost for 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.

Columns/attributes from the above table will be used to come up with new fields as per the formula requirement. New fields that have created to compute this formula has been described as part of Data Preparation section.

As the understanding of the data which is a key step is completed along with the basic details of the data like the attributes, their data types, etc., then next activity is preparing the data.

Chapter 8: Data Preparation

As the data that is available for predicting CLTV in this project is a historic transactional data from a retail apparel firm, there is a need to perform initial exploratory data analysis (EDA) to understand the data fields by using python.

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. Figure No. 8.1 is a snapshot of the top 5 rows of the data after constructing the dataset as needed:

	Customer_ID	Order No	Order Date	Quantity	MRP	TotalSales
0	14955	M011000	2018-01-05 02:08:00	1.0	945	945.0
1	2532	M0110001	2019-08-22 15:12:00	1.0	3095	3095.0
2	10850	M0110002	2019-08-22 16:01:00	1.0	1095	1095.0
3	19829	M011001	2018-01-05 05:16:00	1.0	995	995.0
4	10993	M0110013	2019-08-22 22:49:00	1.0	2995	2995.0

Figure No. 8.1 First 5 rows

Checking for missing values in the data as part of preparing for data modeling. Figure No. 8.2 is a snapshot of the data preparation step to check for missing values based on Customer_ID feature.

	Count	Proportion
Customer_ID	0	0.0
Order No	0	0.0
Order Date	0	0.0
Quantity	0	0.0
MRP	0	0.0
TotalSales	0	0.0

Figure No. 8.2 Missing Values Check

As observed, there are no missing values in the data based on Customer_ID field. Data looks to be clean and ready for the next step. Descriptive statistics on the data is as shown in the Figure No. 8.3. with some of the key statistics of the data

	Customer_ID	Quantity	MRP	TotalSales
count	17782.000000	17782.000000	17782.000000	17782.000000
mean	10263.435778	1.038916	1027.847824	1058.721179
std	6055.332535	0.284685	537.298888	568.953100
min	2.000000	1.000000	95.000000	95.000000
25%	4746.000000	1.000000	595.000000	595.000000
50%	10239.500000	1.000000	995.000000	995.000000
75%	15539.000000	1.000000	1295.000000	1295.000000
max	20610.000000	20.000000	3495.000000	6725.000000

Figure No. 8.3 Descriptive Statistics

Further analysis on the data for details like the time range, total number of unique customers, total quantity sold, etc. Figure No. 8.4 depicts the key metrics.

```
The Time range of transactions is: 2018-01-03 to 2020-12-02
Total number of unique customers: 8381
Total Quantity Sold: 18474.0
Total Sales for the period: 18826180.0
```

Figure No. 8.4 Key Metrics

8.1 Feature Engineering

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.

Below 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

Once the data has been prepared by cleaning and by addition of new features through feature engineering, the data is ready for modeling.

Data preparation is the crux of the steps in the CRISP DM methodology as the data is thoroughly examined and prepared for analysis and modelling. Once the data is prepped, data will be modelled as part of the next chapter.

The features listed here are created from the raw data and then used in the equation to predict the CLTV which is then model using each of the modeling techniques.

Chapter 9: Modeling

Modeling the data to predict CLTV is the agenda of this chapter, the data prepared as part of the previous chapter is now being modeling using various machine learning techniques.

As part of this project, an extensive analysis has been carried out by predicting 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 (Chi Square Automatic Interaction Detection), CART (Classification and Regression Trees), Quest (Quick Unbiased Efficient Statistical Tree) and ExChaid (Exhaustive CHAID). Neural Networks – MLP (Multi-Layer Perceptron) and RBF (Radial Basis Function) and KNN – K Nearest Neighbors

9.1 Linear Modeling Techniques

9.1.1 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.

Figure No. 9.1 shows the total number of transactions considered for prediction of CLTV. Also, it shows the first 5 rows of the CLTV calculations:

```
Total number of transactions happened in the given period: 15908
```

	Name	Sum of Gross Margin	Sum of Product Discounted %	Prop_churn	prop_alive	t	Acquisition cost	Numerator	1+i	(1+i)t	CLTV
0	Priyanka Khandelwal	-4504.690	0.000000	0.00795	0.99205	0.033333	0	651.528838	1.99205	1.023238	636.732428
1	Protima Tiwary	153.000	0.499371	0.00924	0.99076	0.000000	0	650.681630	1.99076	1.000000	650.681630
2	arjita grover	3025.979	4.297736	0.01411	0.98589	3.666667	0	647.483258	1.98589	12.373779	52.327042
3	Swati Gandhi	-3532.410	0.500000	0.03802	0.96198	0.000000	0	631.780365	1.96198	1.000000	631.780365
4	Sonal Somani	365.170	0.400000	0.04787	0.95213	5.300000	0	625.311378	1.95213	34.649446	18.046793

Figure No. 9.1 Top 5 Rows of CLTV

Figure No. 9.2 depicts the box plot of the CLTVs predicted using the Equation 1.1.

```
sns.boxplot(y='CLTV',data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f1642c99090>
```

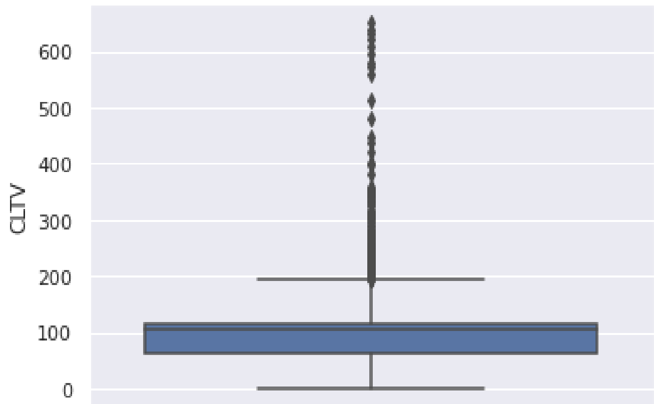

Figure No. 9.2 Boxplot of CLTV values

As the data is modeled using Logistic Regression machine learning algorithm, the CLTVs are being predicted or calculated.

Figure No. 9.3 shows the monthly average of the CLTVs predicted using Logistic Regression:

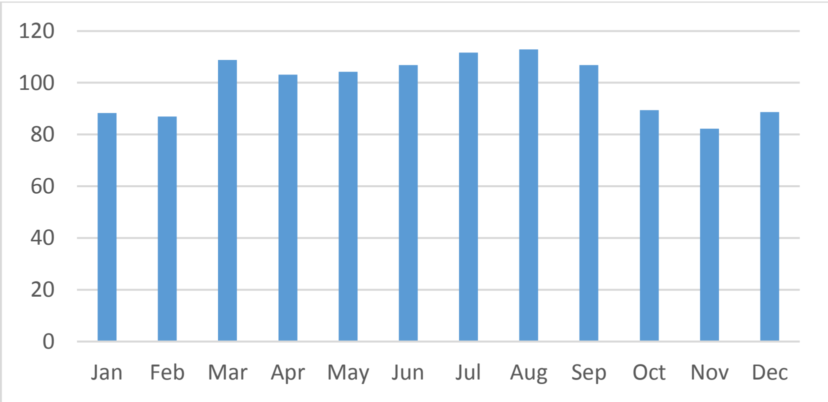

Figure No. 9.3 LR Monthly Average CLTV

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 (Chowdhury & Turin, 2020).

Different variable selection methods part of Logistic Regression have been used to predict the CLTV, the methods used are:

9.1.2 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. Table No. 9.1 and Table No. 9.2 shows the statistics and Figure No. 9.4 and Figure No. 9.5 shows the monthly average CLTVs of these two methods respectively.

Regression Stats	
Multiple R	0.67
R Square	0.45
Adjusted R Square	0.45
Standard Error	402.94
Observations	15907

Table No. 9.1 Backward Conditional Stats

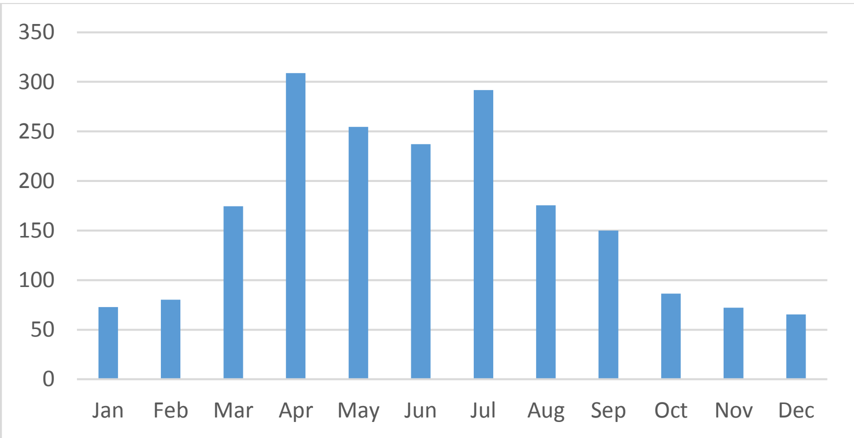

Figure No. 9.4 Backward Conditional Monthly Average CLTV

Regression Stats	
Multiple R	0.67
R Square	0.45
Adjusted R Square	0.45
Standard Error	402.94
Observations	15907

Table No. 9.2 Backward Likelihood Ratio Stats

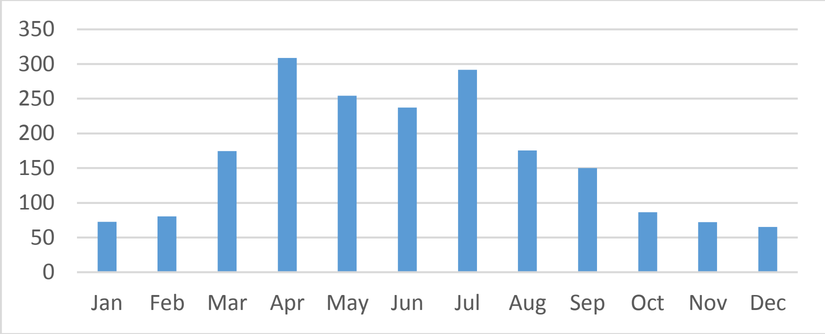

Figure No. 9.5 Backward Likelihood Ratio Monthly Average CLTV

9.1.3 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. Table No. 9.3 and Table No. 9.4 shows the stats and Figure No. 9.6 and Figure No. 9.7 shows the monthly average CLTVs of these two methods:

Regression Stats	
Multiple R	0.68
R Square	0.46
Adjusted R Square	0.46
Standard Error	407.38
Observations	15907

Table No. 9.3 Forward Conditional Stats

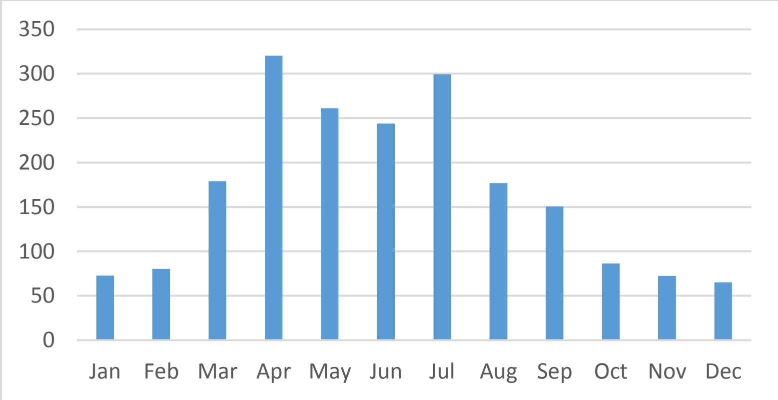

Figure No. 9.6 FC Monthly Average CLTV

Regression Statis	
Multiple R	0.68
R Square	0.46
Adjusted R Square	0.46
Standard Error	407.38
Observations	15907

Table No. 9.4 Forward Likelihood Ratio Stats

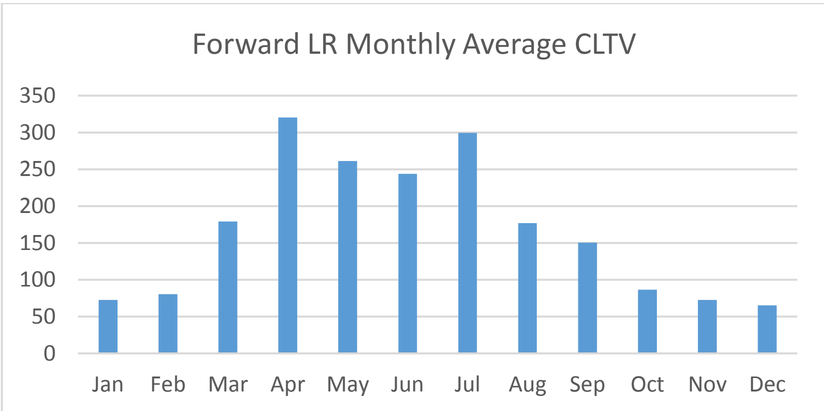

Figure No. 9.7 Forward Likelihood Ratio Monthly Average CLTV

The variable selection methods that are being used here to model the predicted CLTV have different R-squared values. Forward Conditional and Forward Likelihood techniques have a higher R-squared value compared to the Backward Conditional and Backward Likelihood.

From this it is understood 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.

9.2 Non –Linear Modelling Techniques

9.2.1 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.

Node that represent decision and Node that represent Leaf are two nodes in a decision tree. Decisions are made via the nodes that represent decisions and branches get created based on the decisions. Nodes that represent leaf do not contain any branches and are output of those decisions. Figure No. 9.8 shows a graphical representation of decision tree (javatpoint, 2021).

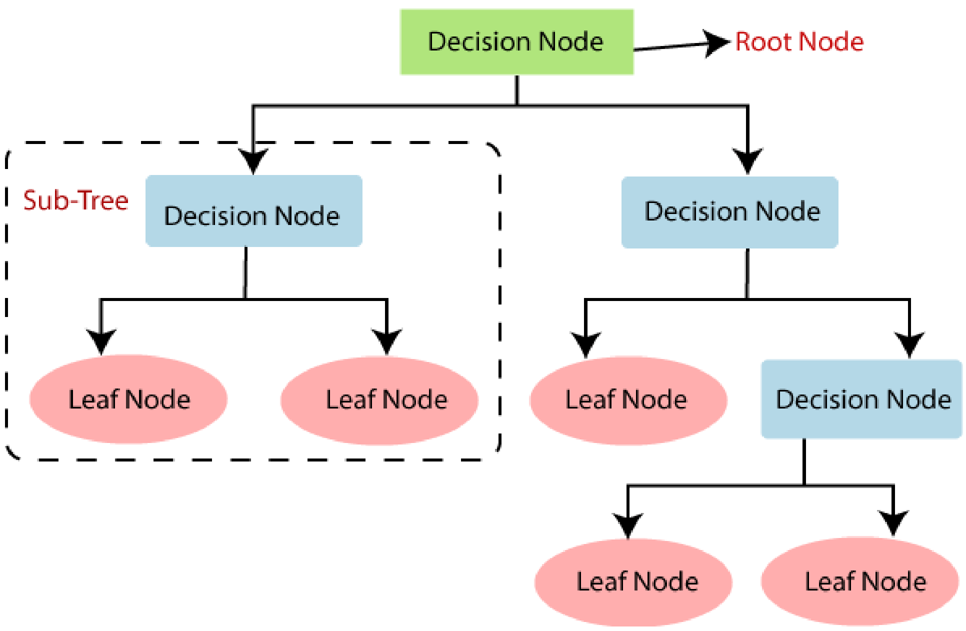

Figure No. 9.8 Decision Tree (javatpoint, 2021)

For this analysis here, these techniques have been made use of:

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.

Table No. 9.5 shows the statistics using CHAID regression, Figure No. 9.9 depicts the CHAID Tree Classifier and Table No. 9.6 shows the confusion matrix along with Figure No. 9.10 showing the monthly average CLTV calculated using CHAID.

Stats	
Standard Error	477.90
Observations	15907

Table No. 9.5 CHAID Stats

CHAID Tree Classifier: Figure No. 9.14 depicts the CHAID Tree Classifier.

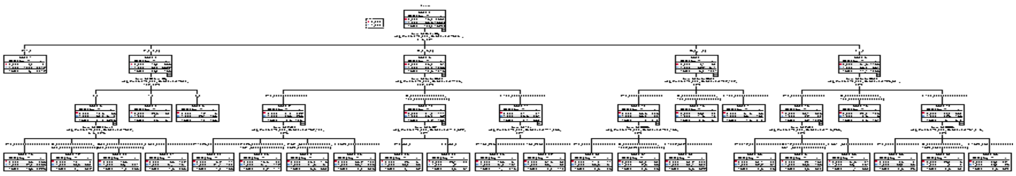

Figure No. 9.9 CHAID Tree Classifier

Confusion Matrix:

	Predicted		
Observed	0	1	Percent Correct
0	1127	1455	43.6%
1	390	12935	97.1%
Overall Percentage	9.5%	90.5%	88.4%

Table No. 9.6 CHAID Confusion Matrix

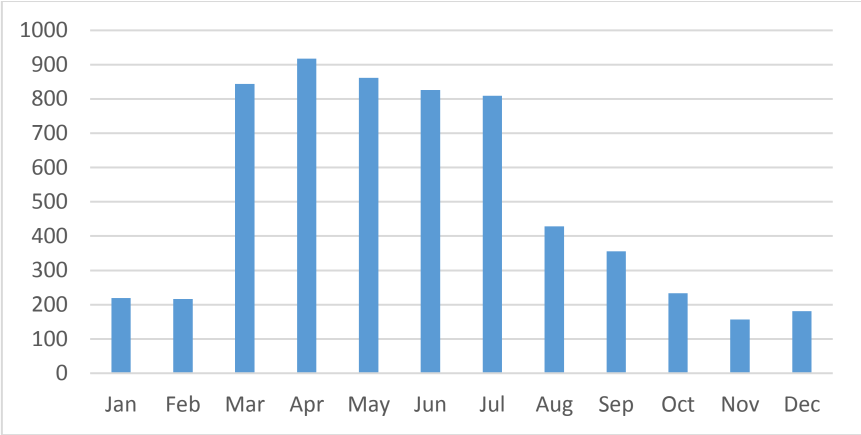

Figure No. 9.10 CHAID Monthly Average CLTV

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. Table No. 9.7 depicts the statistics using CART, Figure No. 9.11 shows the CART tree classifier, Table No. 9.8 shows the confusion matrix and Figure No. 9.12 depicts the monthly average CLTV calculated using CART:

Stats	
Standard Error	362.19
Observations	15907

Table No. 9.7 CART Stats

CART Tree Classifier:

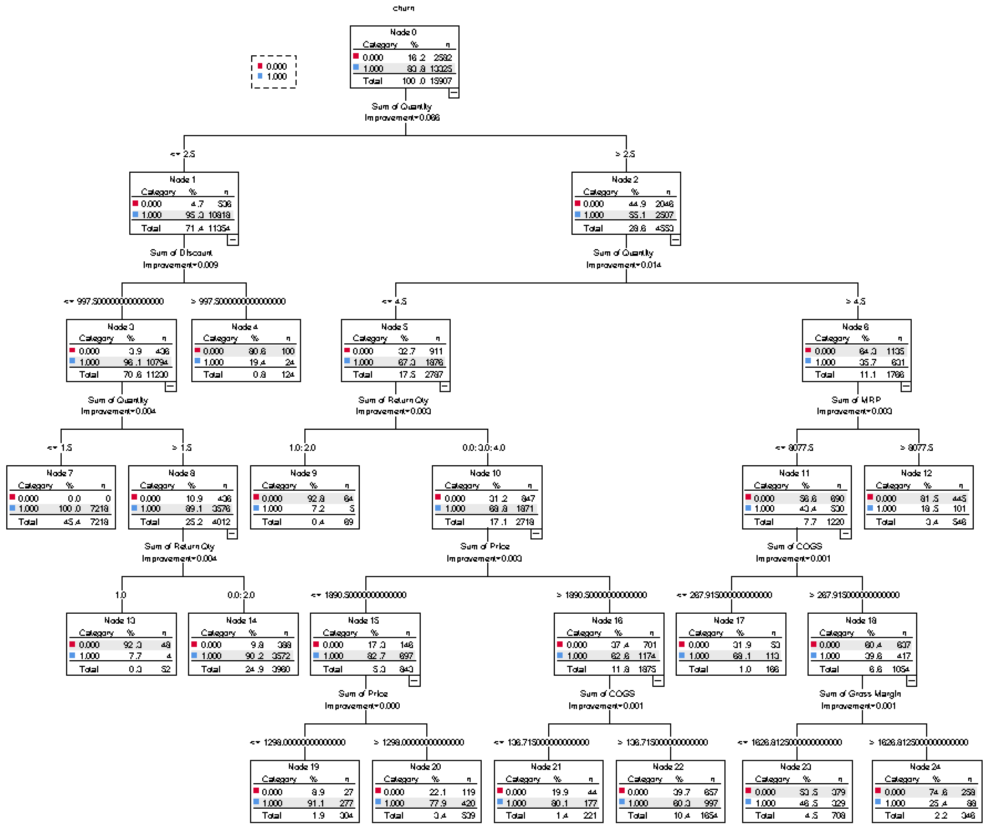

Figure No. 9.11 CART Tree Classifier

Confusion Matrix:

	Predicted		
Observed	0	1	Percent Correct
0	1294	1288	50.1%
1	551	12774	95.9%
Overall Percentage	11.6%	88.4%	88.4%

Table No. 9.8 CART Confusion Matrix

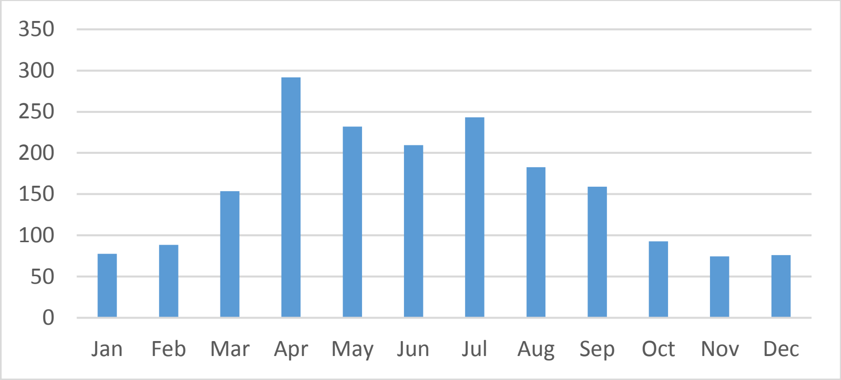

Figure No. 9.12 CART Monthly Average CLTV

Exhaustive CHAID: is a modification to the basic CHAID algorithm that performs a more thorough merging and testing of predictor variables. Table No. 9.9 shows the statistics using ExCHAID, Figure No. 9.13 shows the tree classifier along with Table No. 9.10 showing the confusion matrix and Figure No. 9.14 depicting the monthly average CLTV calculated using ExCHAID:

Regression Stats	
Standard Error	284.45
Observations	15907

Table No. 9.9 ExCHAID Stats

ExCHAID Tree Classifier:

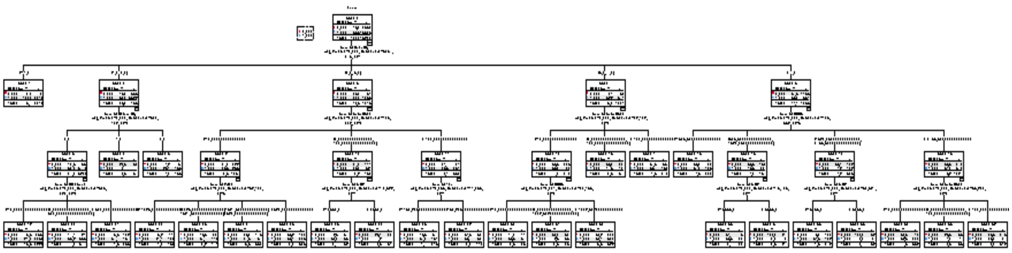

Figure No. 9.13 ExCHAID Tree Classifier

Confusion Matrix:

	Predicted		
Observed	0	1	Percent Correct
0	1179	1403	45.7%
1	430	12895	96.8%
Overall Percentage	10.1%	89.9%	88.5%

Table No. 9.10 ExCHAID Confusion Matrix

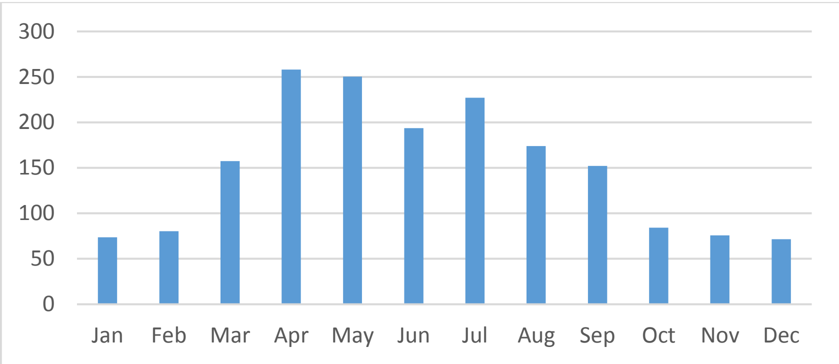

Figure No. 9.14 ExCHAID Monthly Average CLTV

QUEST: stands for Quick, Unbiased and Efficient Statistical Tree. Quest method provides selection of features unbiased and categorical variables are handled with several categories. Table No. 9.11 depicts the statistics using QUEST, Figure No. 9.15 shows the tree classifier, Table No. 9.12 shows the confusion matrix and Figure No. 9.16 shows monthly average CLTV calculated:

Stats	
Standard Error	366.71
Observations	15907

Table No. 9.11 QUEST Stats

QUEST Tree Classifier:

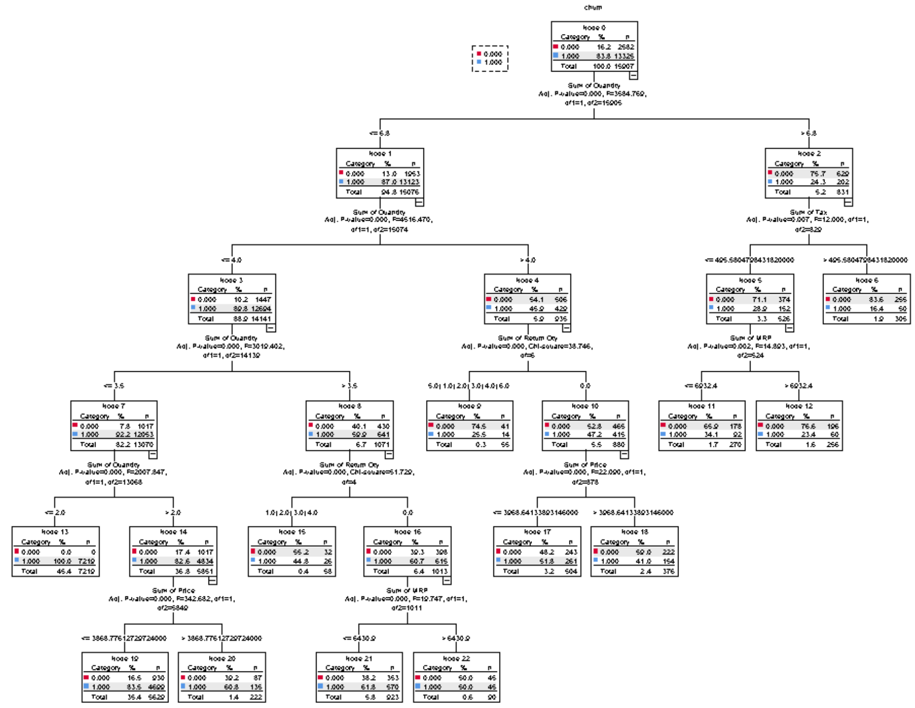

Figure 9.15 QUEST Tree Classifier

Confusion Matrix:

Observed	Predicted 0	Predicted 1	Percent Correct
0	924	1658	35.8%
1	396	12929	97.0%
Overall Percentage	8.3%	91.7%	87.1%

Table No. 9.12 QUEST Confusion Matrix

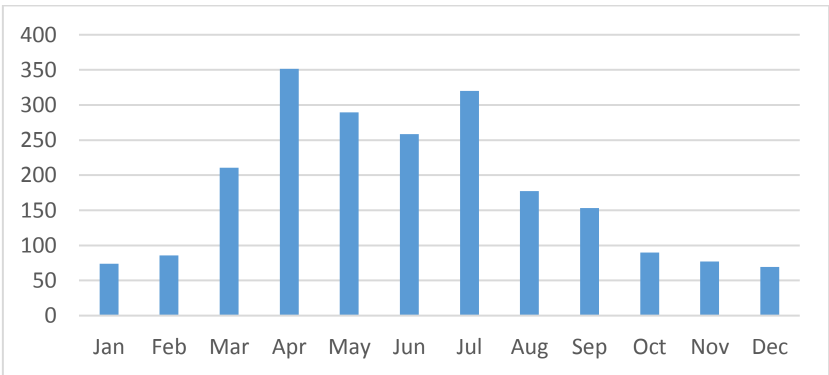

Figure No. 9.16 QUEST Monthly Average CLTV

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.

9.2.2 Neural Networks

Neural Networks are a series of algorithms that mimics the way human brain works in identifying the relationships in a dataset (Pang et al., 2020). Figure 9.17 shows a graphical representation of Simple Neural Network.

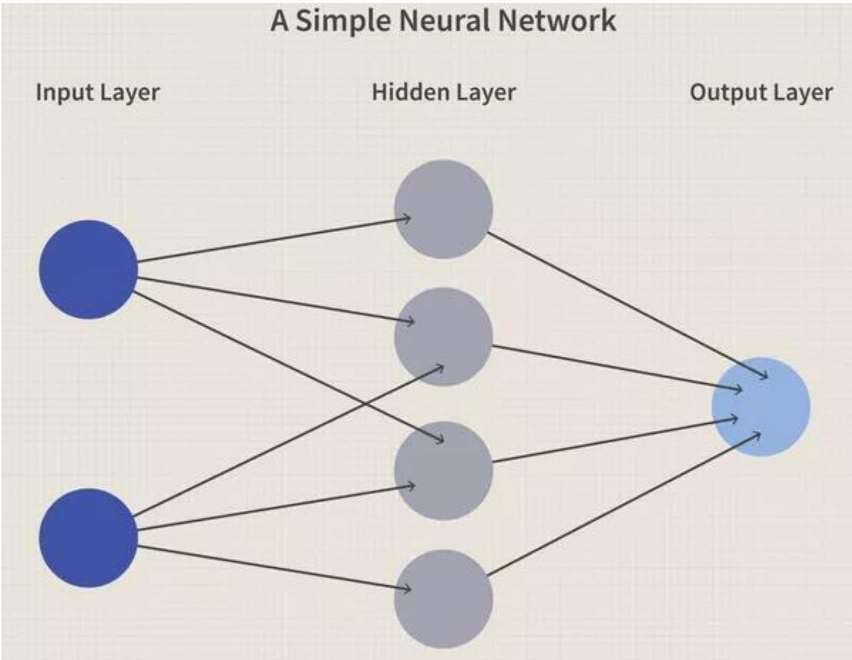

Figure No. 9.17 Simple Neural Network

For the CLTV prediction in this project, these Neural Network techniques have been used:

Multi-Layered Perceptron: In MLP, perceptrons 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. Table No. 9.13 depicts the Statistics for MLP, Figure No. 9.18 depicts the MLP Network derived out of the modelling, and Table No. 9.14 shows the MLP confusion matrix and Figure No. 9.19 represents monthly average CLTV calculated using MLP.

Stats	
Standard Error	358.59
Observations	15907

Table No. 9.13 MLP Stats

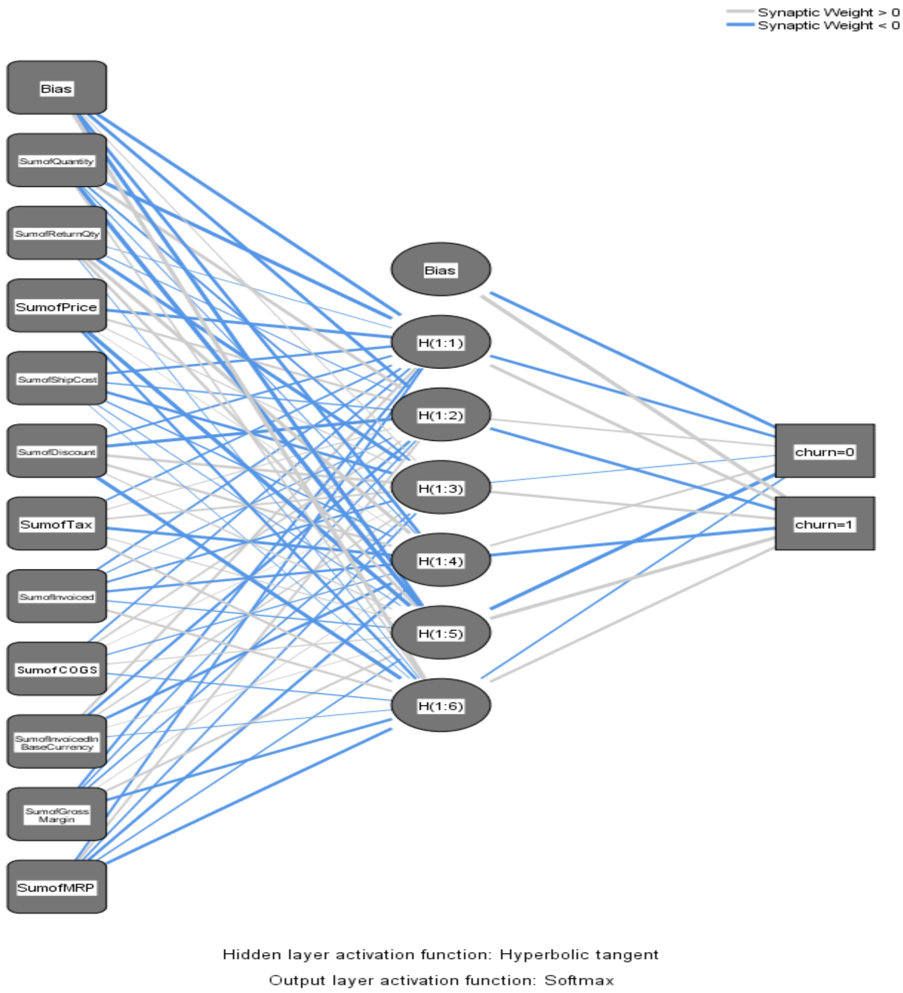

Figure No. 9.18 MLP Network

Confusion Matrix:

Sample	Observed	Predicted 0	Predicted 1	Percent Correct
Training	0	860	949	47.5%
	1	382	9015	95.9%
	Overall Percent	11.1%	88.9%	88.1%
Testing	0	392	381	50.7%
	1	188	3740	95.2%
	Overall Percent	12.3%	87.7%	87.9%

Table No. 9.14 MLP Confusion Matrix

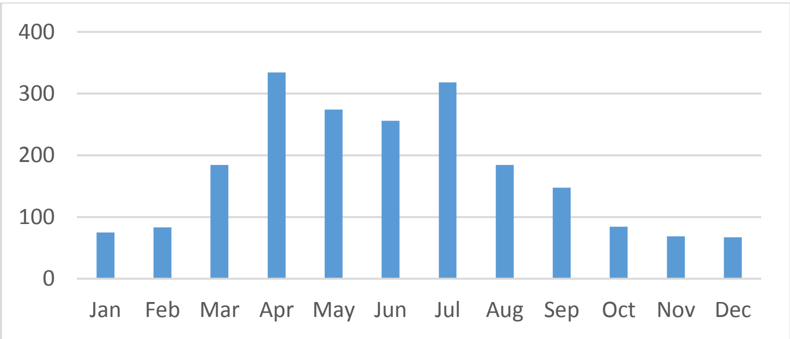

Figure No. 9.19 MLP Monthly Average CLTV

Radial Basis Function: 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. Table No. 9.15 shows the Statistics for RBF and Figure No. 9.20 shows monthly average CLTV.

Stats	
Standard Error	336.35
Observations	15907

Table No. 9.15 RBF Stats

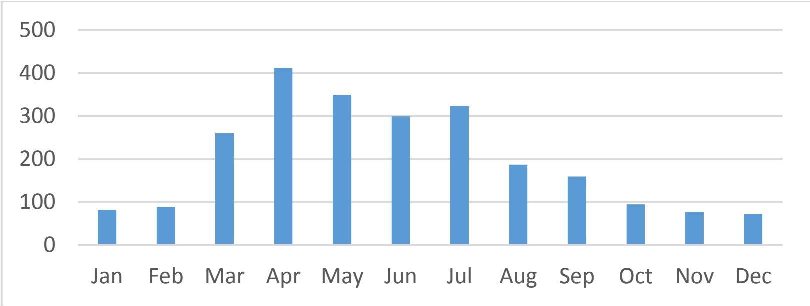

Figure No. 9.20 Radial Basis Function Monthly Average CLTV

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 R-squared value of 0.57 looks to be a better model compared to Multi-Layer Perceptron with a value of 0.51.

9.2.3 K-Nearest Neighbor (KNN)

KNN is a non-parametric, supervised learning, classification algorithm which uses proximity to make classifications or predictions. Table No. 9.16 shows the KNN Stats.

Below is the statistics with R-squared using KNN:

Stats	
Standard Error	310.23
Observations	15905

Table No. 9.16 KNN Stats

Figure No. 9.21 shows the KNN modelling and Figure No. 9.22 shows the monthly average CLTV calculated using KNN.

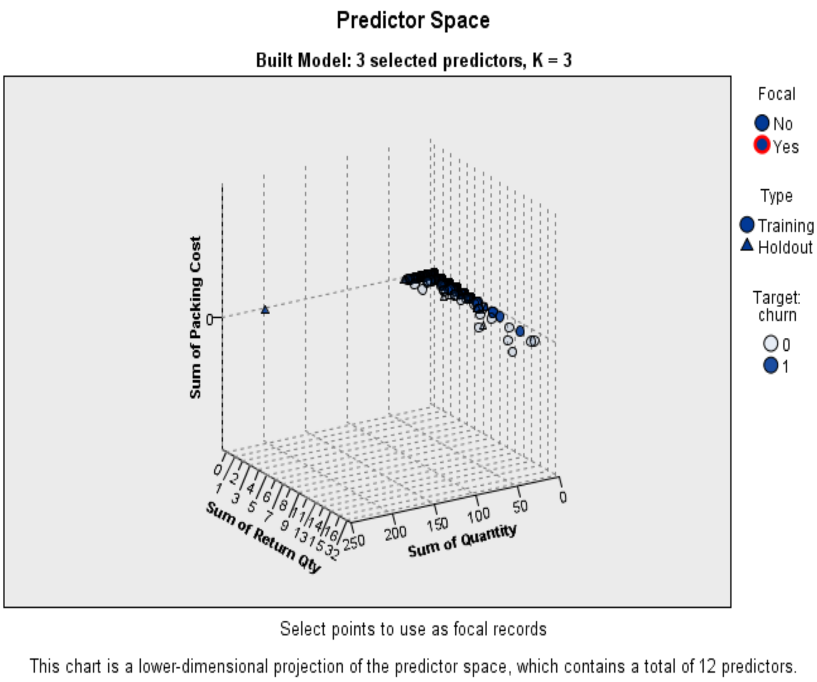

Figure No. 9.21 KNN Predictor Space

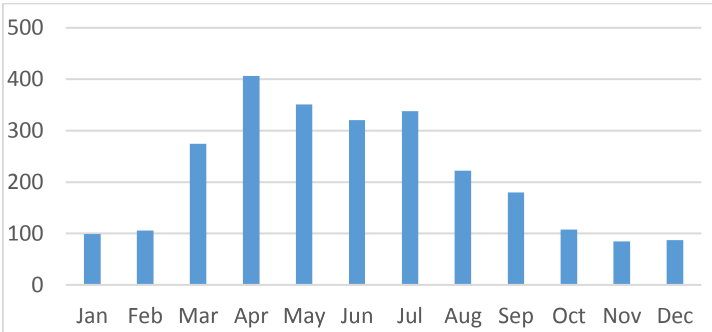

Figure No. 9.22 KNN Monthly Average CLTV

KNN model has a lesser 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. In the next chapter, these models will be evaluated to derive the value to be passed on to the business.

R-squared values generally ranges from 0 to 1. 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 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 (CFI, 2019).

Chapter 10: Model Evaluation

Models built as part of the previous chapter are being evaluated with the results to check on the predicted CLTVs.

CLTV has been calculated or predicted using Linear and Non-Linear machine learning techniques. CLTV thus predicted are used to derive the 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. Figure No. 10.1 shows the comparison of Mean Absolute Error of both linear and non-linear models:

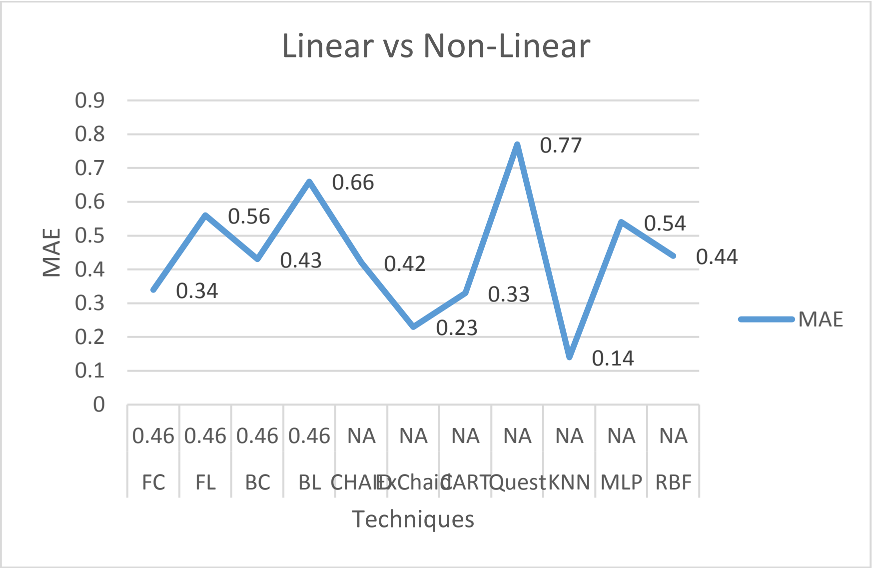

Figure No. 10.1 Comparison of Models

Table 10.1 shows the comparison of key metrics of the different modeling techniques used in this project:

Algorithm	**R Squared**	**Mean Absolute Error**	**MAPE**	**Min. APE**	**Max. APE**	**RMSE**
FC	0.46	0.34	19%	15%	25%	0.9
FL	0.46	0.56	27%	18%	30%	0.6
BC	0.45	0.43	30%	20%	35%	0.01
BL	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

Table No. 10.1 Comparison of metrics

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.

As the evaluation of model is completed, the next chapter talks about the deployment process that needs to be carried out.

Chapter 11: Deployment

This chapter looks like into the deployment options and the details of deploying this solution.

As part of this project work, the modeling has been carried out using the data from csv file with python scripting on the Google Collab. Modelling of the data has been carried out by making use of both Linear and Non-Linear machine learning techniques.

A well planned deployment can be implemented once the business reviews and approves this study. The next chapter looks into analyzing the results of the model developed.

Chapter 12: Analysis and Results

Here in this chapter, the CLTV predicted is analyzed to derive the results. 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.

In the next chapter, this analysis helps with the recommendations and conclusions.

Chapter 13: Conclusions and Recommendations for future work

This chapter concludes this exercise of predicting the CLTV and provides recommendations based on the CLTV. 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.

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.

Customer value can be used to come up business plan 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.

This chapter concludes the project with providing the suggestions based on the predicted CLTVs in order to achieve the objectives.

Bibliography

Burkard, K. (2020). *What is Customer Retention, and Why is it Important?* Smile Blog. https://cs-agents.com/blog/what-is-customer-activation-and-why-is-it-important/

CFI. (2019). *R-Squared - Definition, Interpretation, and How to Calculate*. https://corporatefinanceinstitute.com/resources/knowledge/other/r-squared/

Chowdhury, M. Z. I., & Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. In *Family Medicine and Community Health* (Vol. 8, Issue 1). https://doi.org/10.1136/fmch-2019-000262

Customer Lifetime Value Prediction using Machine Learning | Addepto. (2019). https://addepto.com/customer-lifetime-value-prediction-machine-learning/

Hardie, B. (2006). Modeling Customer Lifetime Value. *Journal of Service Research*, *Volume 9*(Clv), 17. https://doi.org/10.1177/1094670506293810

How Retail Brands Can Predict Customer Lifetime Value. (2019). https://tinuiti.com/blog/ecommerce/customer-lifetime-value/

javatpoint. (2021). *Machine Learning Decision Tree Classification Algorithm - Javatpoint*. Javatpoint.Com. https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

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

Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2020). An innovative neural network approach for stock market prediction. *Journal of Supercomputing*, *76*(3), 2098–2118. https://doi.org/10.1007/S11227-017-2228-Y

Predicting Customer Lifetime Value : A Definitive Guide. (2020). https://www.analyticsvidhya.com/blog/2020/10/a-definitive-guide-for-predicting-customer-lifetime-value-clv/

Sharapa, M. (2019). *5 Simple Ways to Calculate Customer Lifetime Value | by Marie Sharapa | The Startup | Medium*. Medium.Com. https://medium.com/swlh/5-simple-ways-to-calculate-customer-lifetime-value-5f49b1a12723

“5 Simple Ways to Calculate Customer Lifetime Value | by Marie Sharapa | The Startup | Medium.” https://medium.com/swlh/5-simple-ways-to-calculate-customer-lifetime-value-5f49b1a12723 (August 6, 2022).

“Customer Lifetime Value Modelling.” https://dr.lib.iastate.edu/entities/publication/61e87bd9-a2ca-4fce-b7e5-5e196a332d48 (August 6, 2022).

“Customer Lifetime Value Prediction Using Machine Learning | Addepto.” https://addepto.com/customer-lifetime-value-prediction-machine-learning/ (March 23, 2022).

"Logistic Regression — Detailed Overview | by Saishruthi Swaminathan | Towards Data Science." https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc (August 6, 2022).
"Predicting Customer Lifetime Value : A Definitive Guide." https://www.analyticsvidhya.com/blog/2020/10/a-definitive-guide-for-predicting-customer-lifetime-value-clv/ (March 23, 2022).
Sharma, Shreya. 2021. "Customer Lifetime Value Modelling.": 1–31.

Appendix

Plagiarism Report[1]

Customer Life Time Value with Machine Learning

ORIGINALITY REPORT

13%	12%	8%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	mafiadoc.com Internet Source	2%
2	ukcatalogue.oup.com Internet Source	2%
3	www.water.wa.gov.au Internet Source	2%
4	dr.lib.iastate.edu Internet Source	1%
5	Submitted to Mid-America Christian University Student Paper	1%
6	amfa.iau-arak.ac.ir Internet Source	1%
7	researcherslinks.com Internet Source	<1%
8	Submitted to University of Durham Student Paper	<1%
9	ieeexplore.ieee.org Internet Source	<1%

[1] Turnitn report to be attached from the University.

Publication in a Journal

Anand Kumar N, Mithun DJ, Rashmi Agarwal "Predicting Customer Lifetime Value using ML: A Comparative Analysis " 9th International Conference on Business Analytics and Intelligence. vol. Volume 9, no. Clv, p. 17, 2006, doi: 10.1177/1094670506293810.

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.

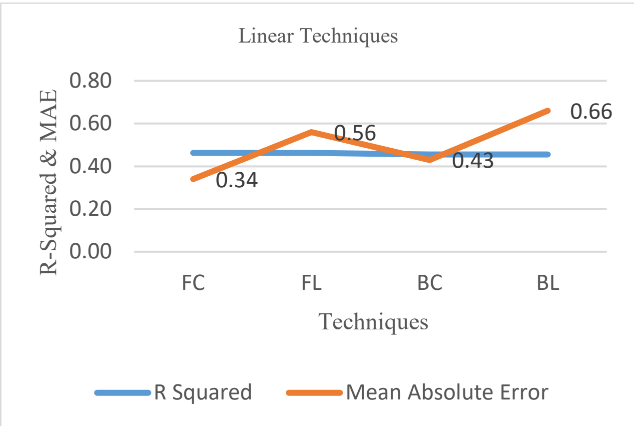

Fig. 1. Linear techniques comparison

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

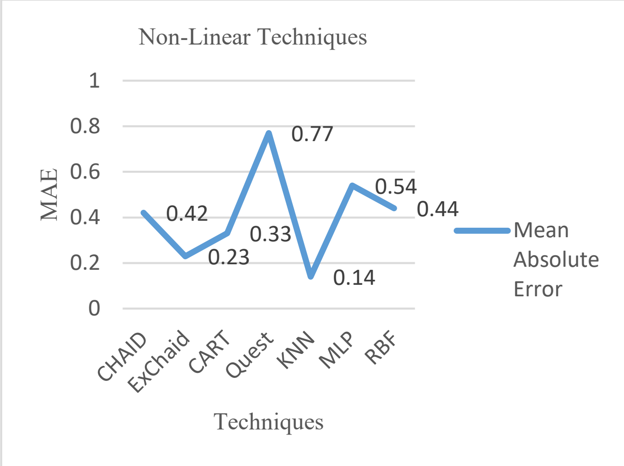

Fig. 2. Non-linear techniques comparison

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).