Customer Lifetime Value Prediction of Motor Insurance Company using Regression Model

SWAYAM PRAKASH JENA[1], THANISH SHEKAR[2], DHIVYA TEFELLA[3], YASHAS B S[4]
[1, 2, 3, 4] *REVA Academy for Corporate Excellence – RACE, REVA University*

Abstract- ***Historically auto insurance companiesput more focus on policy sales as an important guiding metric when it comes to measuring their marketing success. New customers are the lifeline of any growing business. But while sales remain an important result of a successful customer acquisition effort, it is important to make sure that policy sales aren't the only metric used to measure performance. Not all customers purchased insurance are equal. Someone who purchases an inexpensive policy is going to be less valuable for business than someone who purchases an expensive one, and longtime customers will bring in more money than those who buy a one-year policy and do not renew. This concept is called customer lifetime value (CLV). And if a company is not paying attention to it, it is going to wind up overpaying for low-value customers and losing out on high-value customers it might have had. As it turns out, modern companies can analyze their historical data to determine the lifetime value of their customers and determine the factors that can affect the CLV.***

I. INTRODUCTION

Customer lifetime value is a powerful and straightforward measure that synthesizes customer profitability and churn (attrition) risk at individual customer level.For existing customers, customer lifetime value can help companies develop customer loyalty and treatment strategies to maximize customer value.For newly acquired customers, customer lifetime value can help companies develop strategies to grow the right customers.

II. LITERATURE REVIEW

The goal of this paper is to Analyze the Sales of the company and predict the Customer lifetime value. The author uses predictive modelling techniques such as Linear Regression, Random Forest Regressor, Gradient Boosting Regressor, XGBOOST to predict CLV. Based on the above prediction methods, it was found that accuracy of Gradient boosting model was better than other models. Based on the predicted CLV, the author wants to perform customer segmentation using classification algorithms like K-means, Logistic Regression, Naive Bayes. It was found that Naive Bayes model is better suited for segmentation.

However, the author concluded that to increase customer Lifetime Value, we should focus on Effective Communication, Loyalty Program, Retargeting, and It is seen from customer segmentation based on predicted CLV, that about 17% customers contribute to almost 50% of the Value. This is the segment that should be targeted by the marketing team. These customers should be nurtured so that they continue with the company and efforts should be made to increase their CLV.

CLTV has been a mainstay concept in Marketing Management for the past few decades. However, most of the literature on the topic is dedicated to extolling the use of CLTV as a decision-making criterion or considered it in the context of business profitability. It is also discussed for the key role it plays in customer acquisition/retention trade-offs and customer acquisition decisions.[5] It is important to measure CLTV as this can be used as a metric in evaluating marketing decisions. It is important for a firm to have an estimate of the customer's lifetime value when the customer first starts doing business with them, and at each of their subsequent purchases.

III. DATA UNDERSTANDING

3.1. DATA ANALYSIS & INTERPRETATION

The data set contains all the transactions starting from the year 2011, it has 9134 rows and 24 columns.

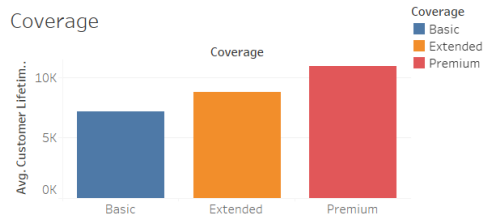
By sourcing proper data andapplying it in the right manner is the key to this research.

3.1.1. DATA UNDERSTANDING (NUMERCAL ATTRIBUTES)

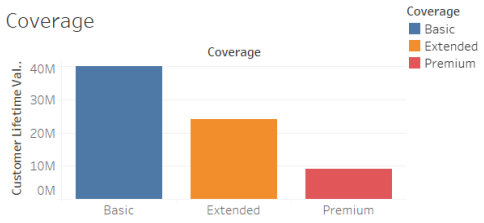
	Customer Lifetime Value	Income	Monthly Premium Auto	Months Since Last Claim
count	9134	9134	9134	9134
mean	8004.94	37657.38	93.21	15.09
std	6870.96	30379.9	34.40	10.07
min	1898.01	0	61	0
25%	3994.25	0	68	6
50%	5780.18	33889.5	83	14
75%	8962.16	62320	109	23
max	83325.38	99981	298	35

	Months Since Policy Inception	Number of Open Complaints	Number of Policies	Total Claim Amount
count	9134	9134	9134	9134
mean	48.06	0.38	2.97	434.08
std	27.91	0.91	2.39	290.50
min	0	0	1	0.09
25%	24	0	1	272.25
50%	48	0	2	383.94
75%	71	0	4	547.51
max	99	5	9	2893.23

3.1.2. DATA UNDERSTANDING (CATEGRICAL ATTRIBUTES)

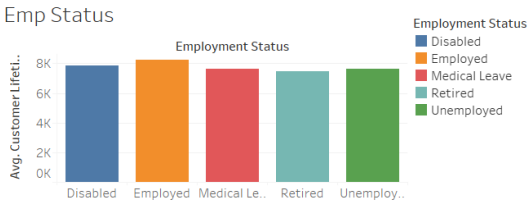


(Fig-1)

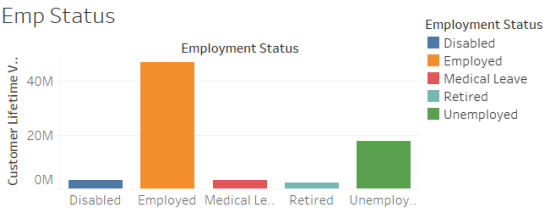


(Fig-2)

People with Premium coverage add more average CLV compared to Basic and Extended coverage. But as the number of customers with Basic coverage is more, they contribute maximum to the total CLV of company.

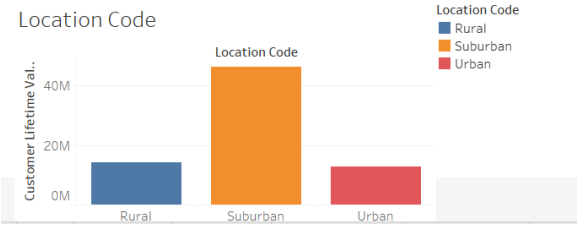


(Fig-3)



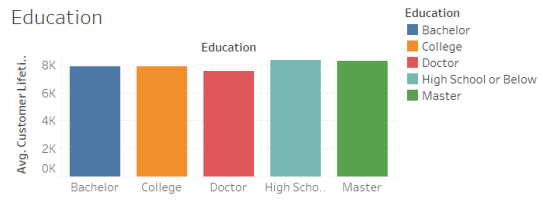
(Fig-4)

Employee status is almost the same across all education sectors for average CLV. But employed customers contribute maximum to the total CLV.

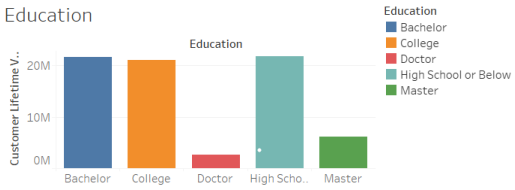


(Fig-5)

Customer lining in suburban contribute maximum to the total CLV.



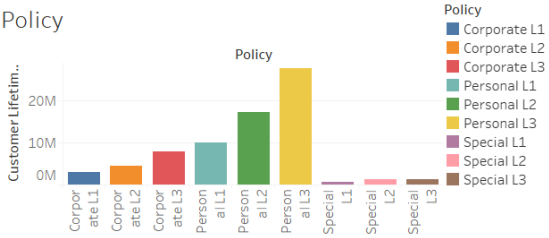
(Fig-6)



(Fig-7)

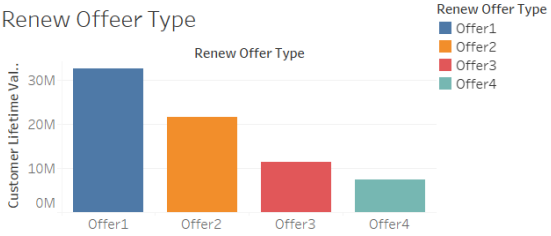
Average CLV does not vary much based on customer education. Whereas customers with masters and doctorate degree education contribute less to total CLV as compared to others.

Based on the analysis, average and total CLV are almost same across male and female customers.



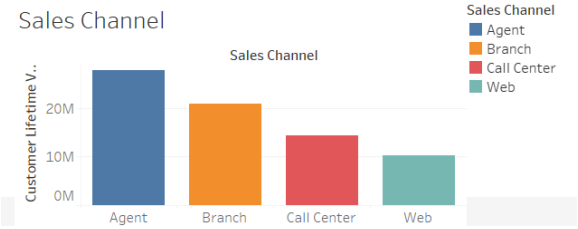
(Fig-14)

Customers with Personal auto policies contribute more to the total CLV.



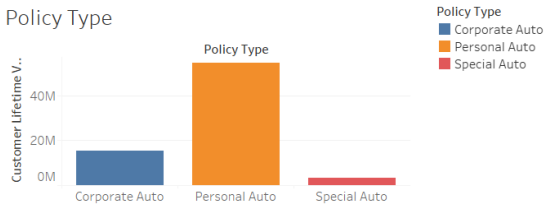
(Fig-16)

When it comes to renewal of the policy, customers have responded more to offer-1 provided by the insurance company.



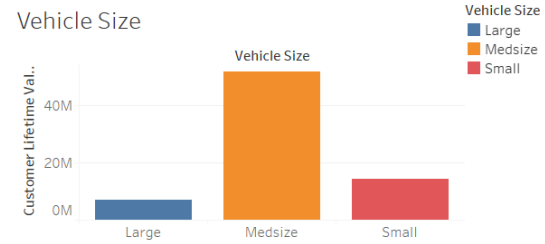
(Fig-18)

If you see the plot, a greater number of policy sales have been done through the agent, Agent has contributed more to the total CLV.



(Fig-20)

Based on the number of policies being sold to the customers, the Personal Auto policy is the most popular policy sold, among others.



(Fig-26)

Based on the analysis, the number of insurance policy being applied are more for midsize vehicles.

3.2. DATA PREPARATION

3.2.1. MISSING VALUE IMPUTATION

There wereno missing valuesfound in the data set.

3.2.2.DATA NORMALIZATION

We used standard scaler to scale the data and fit it into different models. The Standard Scaler is a normalization technique that transforms the features of a dataset to have zero mean and unit variance. The mathematical formula for Standard Scaler normalization is as follows:

For each feature (column) in the dataset:

- Calculate the mean (μ) of the feature.
- Calculate the standard deviation (σ) of the feature.
- Subtract the mean from each value in the feature.
- Divide each value by the standard deviation.

The formula to perform Standard Scaler normalization on a feature x is:

$x\_normalized = (x - \mu) / \sigma$

Where:

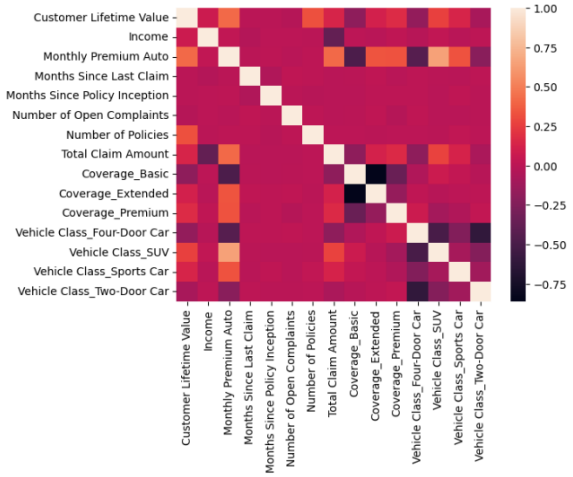
- x_normalized is the normalized value of the feature x.
- x is the original value of the feature.
- μ is the mean of the feature.
- σ is the standard deviation of the feature.

This normalization process ensures that the transformed feature has a mean of zero and a standard deviation of one. It helps in bringing different features to a similar scale and can be beneficial for certain machine learning algorithms that are sensitive to feature scales.

3.2.3.OUTLIER TREATMENT

During analysis, we found that there are considerable number of outliers present in the Columns such as Customer Lifetime Value, Monthly Premium Auto and Total Claim Amountthat may impact the model. So, we have to drop the outliers of these columns before fitting in to the model.

3.2.4.CORRELATION ANALYSIS



(Fig-19)

3.2.5.FEATURE ENGINEERING

Out of 24 variables, based on the analysis we found that 14 variables had a significant impact and can be used as features to predict CLTV. We performed Ttest to check if a variable has significance on Customer Life Time Value or not. If the variable has p-value>0.05 then the variables are rejected.

IV. DATA MODELING

Both industry and research efforts have increased to help shape many different methods for CLTV estimations. Both statistical and machine learning techniques are used extensively for this purpose. Here we used Regressions models to predict CLTV.

Clustering or a multi-class classification problem may not be the most appropriate way since the problem is more of regression problem in which our goal is to estimate a continuous value. In regression models the models are trained to learn and predict continuous values.

We used the below Regression Models to verify results.

LR: Linear Regressor

KN: K Nearest Neighbor Regressor

DT: Decision Tree Regressor

GB: Gradient Boosting Regressor

CLTV is used as the target variable and the data set is divided into train and test sets. Training set is used to train the model and the model is validated on the test set. This helps in understanding the model accuracy. We have

4.1Linear Regression

One of the most well-known algorithms in machine learning is Linear regression. This algorithm fits a linear equation on the observed dataset to model the relationship between two variables. The goal of linear regression is to find the best-fit line that summarizes the relationship between the variables, where the line minimizes the differences between the actual values and the predicted values.

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \varepsilon$

Where:

- y is the dependent variable (the variable you want to predict).
- β_0 is the intercept or bias term.
- β_1, β_2, ..., β_n are the coefficients or weights associated with the independent variables x_1, x_2, ..., x_n respectively.
- x_1, x_2, ..., x_n are the independent variables (also known as features or predictors).
- ε represents the error term or residual, which accounts for the variability in the dependent variable that cannot be explained by the independent variables.

4.2KNeighborsRegressor

KNeighborsRegressor is a type of regression that uses the k-nearest neighbors approach for predicting the value of a continuous target variable.During prediction, the algorithm searches for the k-nearest neighbors of the input point in the feature space based on a distance metric, and then predicts the target value as the mean of the target values of the k-nearest neighbors.

- Data Preparation:

We have a dataset with observations consisting of p independent variables (features) and a corresponding dependent variable (target variable).Each observation is represented as a vector in a p-dimensional feature space.

- Distance Calculation:

KNN uses a distance metric (e.g., Euclidean distance) to measure the similarity between observations in the feature space.

For two points $P = (p_1, p_2, ..., p_n)$ and $Q = (q_1, q_2, ..., q_n)$ in an n-dimensional Euclidean space, the Euclidean distance between them is given by:

$d(P, Q) = \sqrt{((q_1 - p_1)^2 + (q_2 - p_2)^2 + ... + (q_n - p_n)^2)}$

- Finding the K Neighbors:

The K nearest neighbors are the training observations with the smallest distances to the test observation.
K is a user-defined parameter and represents the number of neighbors to consider.

- Predicting the Target Value:

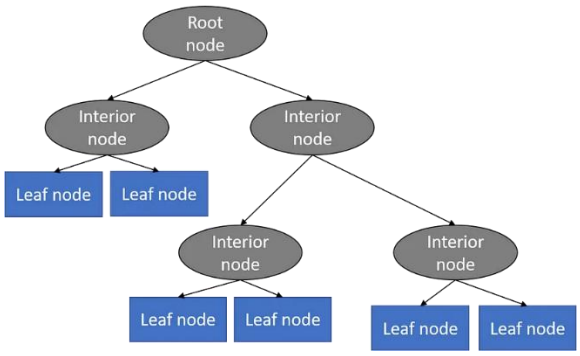
In regression, the predicted target value for the test observation is typically computed as the average (or weighted average) of the target values of its K nearest neighbors.This means that the predicted value is the mean of the dependent variable values of the K nearest training observations.

- Model Evaluation:

The performance of the K-nearest neighbors regressor can be evaluated using various metrics, such as mean squared error (MSE) or R-squared.

4.3 Decision Tree Regressor

Decision Tree Regressor is a type of regression algorithm that uses a decision tree to model the relationship between the input features and the target variable. During training, the algorithm selects the best feature to split the data based on a criterion such as the reduction in variance or the increase in information gain. It continues to split the subsets until it reaches a stopping criterion. During prediction, the algorithm traverses the decision tree from the root node to a leaf node, where it predicts the target value as the mean of the target values of the training samples in that leaf node.



(Fig-20)

4.4Gradient Boosting Regressor

Gradient Boosting Regressor is a type of ensemble learning algorithm used for regression tasks. It combines multiple weak regression models, usually decision trees, to create a strong model that can accurately predict a continuous target variable based on one or more input features. It fits a decision tree to the residuals, or errors, of the previous tree, rather than the target variable directly. This allows it to progressively reduce the errors and improve the accuracy of the model. The residuals are calculated as the difference between the predicted and actual target values.

Ensemble Construction:

a. Initial Prediction:$F_0(x) = \text{argmin}\square \sum_i L(y_i, s)$

$F_0(x)$ represents the initial prediction or base prediction for the target variable y based on the loss function L.

s is a constant that minimizes the sum of the loss function over all training samples.

$L(y_i, s)$ represents the loss function applied to the true target value y_i and the initial prediction s.

b. Gradient Calculation:For each iteration m, calculate the negative gradient of the loss function with respect to the previous prediction:

$g_i\square = -\partial L(y_i, F\square_{-1}(x_i)) / \partial F\square_{-1}(x_i)$

$g_i\square$ represents the negative gradient for the i-th training sample at iteration m.

$F\square_{-1}(x_i)$ represents the prediction made by the ensemble at iteration m-1 for the i-th sample.

c. Fitting Weak Learners:Fit a weak regression model (often a decision tree) to predict the negative gradients ($g_i\square$) obtained in the previous step.

$h\square(x) = \text{WeakLearner}(x, g_i\square)$

$h\square(x)$ represents the prediction made by the weak learner (e.g., decision tree) at iteration m for the input sample x.

WeakLearner denotes the weak learning algorithm used to fit the weak regression model.

d. Ensemble Update:Update the ensemble by adding the prediction of the weak learner scaled by a learning rate (η):

$F\square(x) = F\square_{-1}(x) + \eta * h\square(x)$

$F\square(x)$ represents the updated prediction made by the ensemble at iteration m for the input sample x.

Prediction:The final prediction made by the GradientBoostingRegressor ensemble is the sum of the initial prediction and the predictions of all weak learners, scaled by the learning rate:

$F(x) = F_0(x) + \eta * \sum\square\ h\square(x)$

F(x) represents the final prediction made by the ensemble for the input sample x.

V. MODEL EVALUATION

Accuracy of a machine learning model is measured by checking how good a model can predict with the given data. We calculated the metrics below to determine a suitable model for our data set.

MAE (Mean Absolute Error) is the absolute difference between the true value and the predicted value for each sample, summing up these differences, and finally taking the average by dividing the sum by the total number of samples.

$MAE = (1/n) * \Sigma|y_i - \hat{y}_i|$

Where:

- n is the total number of samples or observations.

- y_i is the true value of the target variable for the i-th sample.
- $\hat{y}_i$ is the predicted value of the target variable for the i-th sample.
- |x| denotes the absolute value of x.

MSE (Mean Squared Error) is the squared difference between the true value and the predicted value for each sample, summing up these squared differences, and finally taking the average by dividing the sum by the total number of samples.

$$MSE = (1/n) * \Sigma(y_i - \hat{y}_i)^2$$

Where:
- n is the total number of samples or observations.
- y_i is the true value of the target variable for the i-th sample.
- $\hat{y}_i$ is the predicted value of the target variable for the i-th sample.

RMSE (Root Mean Squared Error) is the squared difference between the true value and the predicted value for each sample, summing up these squared differences, dividing the sum by the total number of samples, and finally taking the square root of the average.

$$RMSE = \sqrt{(MSE)}$$

MAPE (Mean Absolute Percentage Error) is the absolute percentage difference between the true value and the predicted value for each sample, summing up these absolute percentage differences, and finally taking the average by dividing the sum by the total number of samples. The result is multiplied by 100 to express the error as a percentage.

$$MAPE = (1/n) * \Sigma(|(y_i - \hat{y}_i) / y_i|) * 100$$

Where:
- n is the total number of samples or observations.
- y_i is the true value of the target variable for the i-th sample.
- $\hat{y}_i$ is the predicted value of the target variable for the i-th sample.
- |x| denotes the absolute value of x.

R-2 (R Squared Value) is the proportion of the total variance in the dependent variable that is explained by the regression model (SSR) relative to the total variance in the dependent variable (SST). Subtracting this ratio from 1 provides the R-squared value.

$$R\text{-squared} = 1 - (SSR / SST)$$

Where:
- SSR (Sum of Squared Residuals) represents the sum of the squared differences between the predicted values and the actual values of the dependent variable.

$$SSR = \Sigma(y_i - \hat{y}_i)^2$$

Where:
- y_i is the actual value of the dependent variable for the i-th observation.
- $\hat{y}_i$ is the predicted value of the dependent variable for the i-th observation.
- Σ denotes the summation, which involves summing up the squared differences across all observations.
- SST (Total Sum of Squares) represents the sum of the squared differences between the actual values of the dependent variable and the mean of the dependent variable.

$$SST = \Sigma(y_i - \bar{y})^2$$

Where:
- y_i is the actual value of the dependent variable for the i-th observation.
- $\bar{y}$ is the mean value of the dependent variable.
- Σ denotes the summation, which involves summing up the squared differences across all observations.

Results using 90% of data for training and 10% in testing.

	LR	KN	DT	GB
MAE	0.25350	0.17475	0.03144	0.03272
MSE	0.11005	0.07555	0.00755	0.00611
RMSE	0.33174	0.27486	0.08689	0.07819
MAPE	5.00047	2.69290	0.39201	0.62335
R^2	0.26717	0.44083	0.94973	0.95929

Results using 80% of data for training and 20% in testing.

	LR	KN	DT	GB
MAE	0.25492	0.18063	0.03351	0.03257
MSE	0.11185	0.08127	0.00846	0.00576
RMSE	0.33444	0.28508	0.09200	0.07591
MAPE	4.56053	3.60725	0.77929	0.62300
R^2	0.27314	0.47186	0.94500	0.96255

Results using 65% of data for training and 35% in testing.

	LR	KN	DT	GB
MAE	0.25949	0.19736	0.03477	0.03296
MSE	0.11611	0.09121	0.00882	0.00611
RMSE	0.34075	0.30201	0.09390	0.07815
MAPE	4.41806	3.16629	0.74038	0.66778
R^2	0.25885	0.41781	0.94372	0.96101

(LR: Linear Regressor, KN: KNeighborsRegressor DT: DecisionTreeRegressor, GB: GradientBoostingRegressor)

From the above results it is evident that Gradient Boosting Regressor model predicts better as compared to other models. It has highest accuracy with least mean square error and R^2.

CONCLUSION

For the purpose of this project, Gradient Boosting Regressor model is considered for CLTV prediction based on the prediction accuracy.Some recommendations to increase customer Life Time Value are:

- Attract customers owning luxury cars, currently they contribute least to total CLTV. But each addition/retention of these customers will add 60-50% of CLTV with respect to other customers.
- Company is doing good on suburban areas (60% of total CLTV coming from suburban area), but should focus on advertisement to attract more people on rural and urban area.
- 67% of the customers are purchasing offline (agent/branch), with respect to just 13% online (web). Company should put focus on web to reach more people and optimize resource cost.

REFERENCES

[1] Junxiang Lu, Ph.D. Overland Park, Kansas, Modeling Customer Lifetime Value Using Survival Analysis − An Application in the Telecommunications Industry, Paper 120-28

[2] Ms. Ramamani Venkatakrishna, REVA Academy of Corporate Excellence, Reva University, Bengaluru, India, Mr. Pradeepta Mishra, Director of AI, Lymbyc, LTI, Ms. Sneha P Tiwari, REVA Academy of Corporate Excellence, Reva University, Bengaluru, India, Customer Lifetime Value Prediction and Segmentation using Machine Learning, International Journal of Research in Engineering and Science (IJRES), ISSN (Online): 2320-9364, ISSN (Print): 2320-9356 , www.ijres.org Volume 9 Issue 8 ǁ 2021 ǁ PP. 36-48

[3] Albert Graf, Peter Maas, University of St. Gallen, Customer value from a customer perspective: A comprehensive review, Journal für Betriebswirtschaft 58(1):1-20(April2008), https://www.researchgate.net/publication/225998618