CUSTOMER LIFETIME VALUE PREDICTION AND SEGMENTATION

Prakhar Kant Kushwaha*1, Raman Kumar*2, Prof. Merin Meleet*3

*1,2,3Information Science And Engineering, Rv College Of Engineering, Bengaluru, India.

ABSTRACT

Customer Lifetime Value (CLV) is the total monetary value of transactions made by customers with the company throughout their lifetime. Modern customer lifetime value models have been able to predict the future relationship between individual customers and companies. With the advancement of predictive modelling, algorithms for estimating CLV have become increasingly popular and even more powerful. The popularity of CLV prediction will increase in the coming time. Clients are divided into revenue segments ranging from highly valued clients to low value clients according to revenue they are expected to generate. In this work, the linear regression ML model is used to predict CLV, and then the K Means clustering is used to divide customers into revenue segments which will help organizations to distinguish different customers based on the profitability. After predicting the CLV, organizations can concentrate on deciding how much to spend on advertising, which customers to target for advertising, and how to plan for a customer's transfer from one market segment to another.

Keywords: Customer Lifetime Value, Linear Regression, K-Means Clustering, Customer's Segmentation.

I. INTRODUCTION

Customer’s lifetime value systems (CLTV) can already predict future relationships between individual clients and companies. With the advance of prediction modelling, the algorithm to estimate CLTV is increasingly and more powerful. Both research and industry efforts helped to form a large system that can earn millions of users daily. CLTV estimates that the innumerable companies and cases of use of products, so there is only an increase in interest in the Estimation models of CLTV. ROI maximizes and strict compliance with the concept of retaining customers only with high CLV is more profitable, leading to a client, but leads to contraction. If this is reflected in decreasing the market share, the financial market is expected to be an adverse reaction. A debate that supports smaller income customers can be dangerous to win some great customers. This is done with problems issued when talking about CLV and EC. This tends to consider the expected value of these measures. It should be considered that it must be distributed to move towards the quantification of risks associated with a particular client. Most CLV applications focused on the service industry, which is a significant context that covers and retention. Customer Value or Customer Life Value (CLV) is the total monetary value of transactions / purchases made by customers with your company throughout their entire life cycle. Here, lifecycle refers to the length of time a customer buys from you before turning to a competitor. Different services to be offered to different segments of customers are identified by the service token numbers. Here, we develop a model to predict the customer lifetime value and create customer segments according to the revenue they will generate predicted by the model and then we assign the segment number to each customer id. This prediction of the future customer lifetime value (CLTV) from past purchase behavior is done using machine learning. The optimal number of segments from the CLTV are calculated and then individual customers are divided into these segments.

II. LITERATURE SURVEY

In “Getting the Most Out of All Your Customers” [17], Reinartz, and V. Kumar found that the revenue from the highest 30% of consumers supported the CLV model was 33% above the highest 30% selected supported the RFM model. Venkatesan and Kumar also compared several competing models for customer selection. Using data on almost 2,000 customers from a business-to-business (B2B) manufacturer, they found that the profit generated from the highest 5% customers as selected by the CLV model was 10% to 50% above the profit generated from the highest 5% customers from other models (e.g., RFM, past value, etc.).

In “Customer-Base Analysis with Discrete-Time Transaction Data” [18], Fader, Peter S., Bruce G. S. Hardie, and Paul D. Berger and “RFM and CLV: Using IsoCLV Curves for Customer Base Analysis,” they showed how RFM variables are often wanting to build a CLV model that overcomes many of its limitations. In the paper they
demonstrated that the Recency, Frequency and Monetary values are sufficient statistics for the CLV model. They came up with an iso-CLV graph which shows the recency, frequency and monetary values that give the same CLV.

In "Recapturing Lost Customers" [19], Blattberg, Thomas and Fox showed the significance of linking acquisition and retention decisions. It was observed that ignoring this link can lead to prediction deviations of up to 6% to 52% from the model. Though low prices had high acquisition probability, it created short duration relations. Thus, it might not be of best interest to retain customers who would want to start the relations again. The authors verified this across two industries empirically and also, they discovered that the businesses should acquire on the basis of their profitability rather than the acquisition and retention costs.

In “Customer Acquisition Promotions and Customer Asset Value” [20], Lewis showed that the promotions which increase the number of customers acquired, might be detrimental in the long term. He observed that the probability of renewal on offering regular newspaper subscription was 70% which decreased to 35% on giving 1 dollar discount. Likewise, renewal probability reduced from 40% to 25% on acquiring customers through 10 dollar discount than the regular subscription acquisition. On an average, offering discount of 35% on acquisition resulted in approximately half the lifetime value than regular price offerings. The statements are according to the effects of promotion in the scanner data in long-term.

In "A Dynamic Programming Approach to Customer Relationship Pricing" [21], Lewis used the approach of dynamic programming to examine the pricing dynamics and discovered that for the existing customers, sensitivity of price is reduced with time and for newly acquired customers, it increases with the amount of time lapsed. Thus, the optimal pricing technique would be to offer discounts based on time lapsed in the business relations like giving 1 dollar discount per week for new customers, 2 dollars on first renewal, 2.5 dollars on second renewal and so on than offering one large discount.

In "The Loyalty Effect" [22], Reichheld, Frederick F., and “Zero Defections: Quality Comes to Services” research work, Reichheld and Sasser found that if the retention of customers was increased by 5%, the profitability of the firm could rise by 25% to 85%. Their work increased the interest in the study of retaining customers and the customer’s loyalty. Reichheld put forward the significance of retaining customers. However, Reinartz and Kumar were against this result and proposed that “the customer lifetime value is driven by the revenue and not the duration of a customer’s business relationship”. Also, Reinartz and Kumar contradicted the results on the basis of what they found in research, the low to medium correlation (.2 to .45) between customer duration and profitability across the four data sets. However, if the relationship between customer loyalty and profitability is non linear, the correlation can be weak.

In “Bagging and Boosting Classification Trees to Predict Churn” [23], Aurélie and Christophe Croux, Lemmens and Croux predicted churn for a U.S. wireless customer database using bagging and boosting technique. In the process of bagging, a binary choice model is estimated sequentially from calibration sample resampled versions. In machine learning, it is also called base classifier. Using aggregation of the group formed by the obtained classifiers, the derivation of final choice model is done. The scheme of sampling is different in boosting compared to bagging. In the process of boosting, classifier is sequentially estimated to the versions of initial samples of calibration reweighted adaptively. The scheme of weighting grants incremented weight in the next iteration to the customers which were misclassified. Due to this, the methods of classification are forced to focus customers whose classification is difficult. From these methods, Lemmens and Croux made the comparison of results with the binary logit model and discovered the comparative increase in predicting more than 16% for the gini coefficient and 26% for the top-decile lift. Using considerable assumptions, they demonstrated that these differences can be worth more than $3 million to the company. This is go together with the conclusions of Neslin who found that the methods of prediction matter and the profit can surge by hundreds of thousands of dollars.

III. ARCHITECTURE

In the first stage, the objectives of the project were identified i.e. to predict CLTV to perform customer segmentation based on their future revenues. Then in the next stage, the type of problem was identified to be a regression problem. For model preparation, the features needed to be selected and data preparation had to be
carried out. For that, data visualization was performed to analyze the data and features were identified to be the sum, count and average of the data belonging to five periods of three months in the third stage. The target variable was identified to be CLV for three months. Thus the linear regression model was prepared after train test split in the fourth stage. In fifth stage, the CLTVs are predicted by different models and linear regression was selected. and merged with the Customer IDs. Then, the segmentation of the customers was done to 2(high value), 1(mid value) or 0(low value).

IV. PROPOSED APPROACH

The data was preprocessed to be fed to the model like removal of transactions without customer ID, with negative prices and quantities, duplicate description was checked. The data for the last month i.e. December 2011 was not available, so transactions of that period were dropped. The data wrangling was performed to analyse the data like the behaviour of repeat customers. The transactions were divided according to the period of purchase into five clusters namely M1, M2, M3, M4 and M5. The data was pivoted on the index of Customer ID and for each of the period, sum, count and average of the data prepared. Sum, count and Average of the five periods were the features for machine learning model. The revenue of the customers in the period M1 was made the target variable for Machine learning prediction. The linear regression model was fitted on the data divided into train to test in ratio 65:35. The r square of the linear regression model was obtained and then the segments were created on the CLV values predicted. So then there were customers segmented into clusters 0,1 and 2.

Dataset Pre-processing

The dataset is processed to change it such that it can be fed to the machine learning model for prediction of customer lifetime value.

Null values: There were NULL records in columns Customer Id and Description. Since the plan was to predict CLTV, the records for each customer ID were required to proceed calculations. So the rows containing NULL values were dropped.
Rows with negative quantities and negative prices: The table entries with negative values were identified. In the quantities column, such rows were removed. Then the entries with negative price of the product had to be removed.

Transactions without Customer ID: The Number of missing values in 'Customer ID' column was found to be 135080 (24%). As the final result has to be the segmented customers identified by their customer IDs, these transactions were dropped from the dataset.

Incomplete data: Period of time for the purchase records is from December 1, 2010 to December 9, 2011. The data for the last month is incomplete. Because we are planning to predict CLTV for the next 3 months and we will be aggregating data monthly, let us ignore the records for the incomplete month.

Calculation of total sales: A column for the total sales was created by taking the product of unit price and quantities of item purchased for each transaction; this was required for creating the model.

Duplicates in the dataset: The dataset is checked for the duplicates, there were no duplicates found in the description column.

Feature Engineering: To build the model, firstly the data was sliced into chunks with three month's data each and the last chunk was taken as the target for predictions. Then we obtained the data frame with Customer ID, Invoice Date, Sales Sum, Sales average and Sales count. The data was divided into $M_1, M_2, M_3, M_4$ and $M_5$ where the ends of respective groups were marked by the dates:

- 2011-12-31: $M_1$
- 2011-09-30: $M_2$
- 2011-06-30: $M_3$
- 2011-03-31: $M_4$
- 2010-12-31: $M_5$

V. RESULT & ANALYSIS

The details of the transactions and their analysis was done. So an approximately one year data record of 25,900 purchases for 4,372 unique users is made in 38 countries (mostly in the UK) from 2010-12-01 to 2011-12-09. Graph was plotted for the transactions against the country, and it was found that most of the users are from the United Kingdom. On plotting the count of customers against the quantity bought by them, it is found that mostly customers buy less than 50 items per order. The 1.96% of records were with negative quantities. There were 10,624 rows with negative quantity values, which is about 2% out of the total number of records in the data frame. In analysis, those repeat customers were considered who made at least two purchases. As can be seen, repeat customers tended to make about 12 purchases or less and the majority of them made a purchase every 12 to 50 days.

The feature used for prediction and their coefficients after fitting on the training is shown in the figure 5.5. The train to test ratio on the dataset is 65:35. R-squared ($r^2_{\text{score}}$) for the linear regression was obtained to be: R-Squared for Train set was found to be 0.71 and for the validation set 0.71. R-Squared is the same for the train and test sets, therefore, there was no overfitting or underfitting of the model. The median absolute error for Train set came out to be 202.41 and for the validation set 205.20. On applying the Kmeans clustering, ranging the number of segments of Customer Lifetime values from 1 to 10, the plot obtained is shown in figure 5.6. The elbow is at number of clusters = 2 or 3. The segmentation was carried further with the number of clusters = 3.
The plot of actual revenue of three months versus predicted values by the model for the training set is shown below. Most of the transactions are of the cost less than 10,000. Other points are scattered around the 45 degrees line. The number of transactions for the testing set were comparatively low and their plot also shows the points under the purchasing cost of 10,000.
VI. CONCLUSION

Customer Lifetime Value is a metric to evaluate the gain from investments in marketing. Using the transaction history of customers, the revenue from different customers was calculated, the customer lifetime value was predicted for the customers on the Online Retail dataset from UCI Machine Learning Repository. The transactions were divided into chunks of three months periods and the sum, count and average of these chunks were used as the features for the model preparation. The prediction was carried out using Linear Regression and r square obtained was 0.71 for both the training and testing data implying no underfitting/overfitting present in the model. The segmentation of customers on their lifetime values was achieved by k-means clustering. Based on the data analysis, we found that the repeat customers tend to make about 12 purchases or less within a year and the majority of repeat customers tend to make a purchase every 12 to 50 days.

We predicted 3-month CLTV for customers of the online retail using linear regression. R-squared value for the test set is 0.71, which is not great but it is a good benchmark to try other regression models such as Epsilon-Support Vector Regression and Random Forest Regressor.

By knowing CLTV, we can develop positive ROI strategies and make decisions about how much money to invest in acquiring new customers and retaining existing ones.

The r square score for the prediction using deep neural networks was < 0.7. The probability models are capable of achieving higher accuracies of >0.9. The accuracy of prediction could be further improved by addition of more features to the model. The addition of RFM score directly to the model features led to decline of the r square value for the prediction using deep neural networks was < 0.7. The accuracy of prediction could be further improved by addition of more features to the model.

VII. REFERENCES