ABSTRACT

Customer Lifetime Value ("CLV") modeling focuses on measuring and understanding the value of a customer based on past customer transaction data. Reliable models that accurately predict CLV is a critical tool that helps a company understand customer purchasing behavior, enabling managers to make focused marketing decisions.

Two common "buy-till-you-die" probability models are used to predict CLV on non-contractual customer retail environments: the Pareto-NBD model (Reference 1), and the BG-NBD model (Reference 2) model. One of the primary benefits of these types of models is that they work with aggregate data and are easy to implement.

INTRODUCTION

The retail sponsor we worked with is a seasonal business with customers purchasing at a higher frequency and order value during the peak season in months 11 and 12. One of the downsides to traditional "buy-till-you-die" models for a seasonal business such as the retail sponsor is the assumption that purchasing rates occur evenly throughout time.

We were tasked to explore how to implement seasonal purchasing elements into these traditional CLV models. In addition, being able to account for the COVID has meaningful implications for forecasting expected customer behavior.

METHODS AND MATERIALS

We utilized the CLVTools package in R to incorporate time-varying covariates into the traditional Pareto-NBD model to perform our analysis.

The retail sponsor provided transactions from a sample of customers for a 5-year period. The dataset included purchase date, total sales amount, and purchase channel for each customer. We trained our model using the first 4 years of data and tested the accuracy of CLV predictions to 3-year and 5-year time horizons.

As the CLV model did a better job of retaining customer value in terms of probability of alive prediction, especially for seasonal customers - customers are predicted to stay alive until the next seasonal period.

RESULTS

The traditional Pareto-NBD model did a good job of predicting overall transaction rates at an aggregate level over a wide time horizon, especially for less seasonal customers.

Adding covariates into the Pareto-NBD model for highly seasonal customers had benefits in terms of accuracy of prediction and was an overall better fit as seen in the graphs below.

CONCLUSION

The covariate Pareto-NBD model accomplished the goal of improving CLV predictions for seasonal customers in both shorter and longer time horizons. While we cannot predict such transformative events like COVID, the covariate model neutralized significant penalization for the drop-off in year 4 purchasing volume in future customer forecasts.

We were able to enhance CLV predictions and uncover insights on customer purchasing behavior with just a couple columns of data. The results of this covariate model can be used as a mechanism to drive and measure the impact of business decisions.

BUSINESS IMPLICATIONS

After completing the data modeling process and implementing an improved CLV model over traditional methods, we connected the results of 5-year CLV predictions to practical business application. Analyzing a customer’s CLV as of a single point in time (e.g., year 5) does not tell the entire story. Two customers with similar forecasts for CLV may have different growth trajectories.

To combat this limitation, we used the improved model to calculate 5-year CLV predictions pre and post pandemic to incorporate the change in CLV into our analysis. By controlling for the impact of COVID in the covariate model, customers that returned to pre-pandemic purchasing volumes also returned to similar levels of CLV. This was a novel way to understand changes in purchasing rates.

REFERENCES