**A Customer Lifetime Value Predictive Analysis for Google Merchandise Store**

Hi Aditya,

I hope you're doing well and a happy new year to you! I was delighted to learn during the interview that you were interested in knowing more about my capstone project with Google. This write-up is an overview of our efforts to develop a predictive model for their merchandise store that will help them in acquiring long-term and valuable customers.

The Google Merchandise Store is an online store that sells Google-branded items. They use digital advertising to drive customers to their website and increase sales. Our group met with a Google representative to discuss their needs. The team's primary goal was to acquire new customers through digital advertising. They aimed to attract customers who would be the most valuable in the long run. They requested that we base our analysis on the customer's value during the first 90 days following their initial purchase and they provided us with their transactional dataset along with their initial model. Our task was to build a model that predicted the long-term value of first-time buyers.

The Google team's initial model used very few transactional-level factors for their analysis, such as the number of pages users viewed on the website, the amount of time they spent on the site, etc. To predict the lifetime value of new customers, it is critical to consider factors such as the channels through which they were acquired, personal characteristics such as their location, and behavioural characteristics such as clicking on product details, adding, or removing items from cart. During our data exploration phase, we discovered that our massive dataset had 90% missing values, resulting in a very low correlation value for most of the characteristics with revenue. However, utilizing our marketing acumen, we chose to include additional variables to gain insights about these consumer’s engagement patterns in addition to the transaction-level data. We divided our dataset into a training set and a test set to account for the dataset's complexity and to effectively evaluate our model's performance.

We started with linear regression because our dependent variable ‘future revenue’ is a continuous variable, but because linear regression is sensitive to outliers, it did not perform well and was therefore unreliable. The decision tree model could not produce correct results, so we attempted utilizing random forest, which combines several decision trees to produce the final result rather than depending just on one decision tree. Although the results of this model showed enhanced accuracy, we chose to test XGBoost Regression instead because we could not rely on the model's independent interpretability since it merely computes an average over numerous decision trees. To manage the behaviour of our XGBoost Regressor model and increase its predictive accuracy, we employed hyperparameter tuning to get the most accurate findings.

Using our improved Customer Lifetime value prediction, the Google team can now provide the system with data inputs that reflect each customer's eventual impact and thus achieve precise targeting for a potential high value shopper. With this insight, budgets can be allocated to the channels that maximize profit over a longer time period which, in our case, was Direct Channel. Google Ads platform can be a great aid in improving the channel further. Overall CLV based model provided more valuable insights compared to return-on-investment model for Google Ads which may not be able to maximize the potential of the ads. With all the insights, techniques and systems put together our model was flexible to perform and adapt to suit multiple client needs.

This project taught us a lot about predicting lifetime value and conducting research as a whole. It gave us hands-on Big Query ML experience, which we used to improve our model. We were able to understand major drivers for improving the algorithm for the store by putting our model to work within the Google smart bidding business challenge.