BLOG POST

In Direct Response Television, the goal is, clearly, to elicit responses. This could mean something different for many clients, maybe a phone call, a sold policy, a free quote. However, for one of our clients, the end goal is a **paid sale**. Direct Response gives us the tools to track someone from just a phone call, to a lead, to a sale. It’s important for us to know exactly what pieces of our advertising are driving response, so we can continue to spend in places that generate the most efficient sales. The problem? Sales take weeks to come in, and we need to buy media today! The solution? Predictive modeling.

In this situation, a decision tree regressor is our chosen method. Television ads have many categorical features: Access Code, Station Code, Length, Program, Affiliate, Market, and Air Type. All of these should be explored as features of our model to see which factors of the ad buy have a strong predictive impact on the resulting cost-per-sale.

Through testing, we have determined that of the 327 different columns available, spend, lead volume, and just a few creatives are considered **feature importances**. This means they will be included in the decision tree model. We begin at the top of the tree, and make a split (True or False) based on what will produce the lowest error. From here, we continue to make splits to minimize error. Based on tests, we have limited the size of the tree to stop at a depth of 5 meaning it splits 4 times before we get to the final results.

After training the model and testing on a held-out dataset, our root mean squared error was 1113. A benchmark to compare this to would be if we had just been predicting the mean cost-per-sale for every record (this is called a null model). The root mean squared error of this null model would be 1570, so our decision tree beat the null model in predicting cost-per-sale.

There are many advantages of a decision tree. First, we can display it graphically and it is highly interpretable. This will be important as we implement this model for media buyers to use in placing their ad buys. Decision trees also outperform linear models when the relationship is non-linear, which is certainly the case in this dataset. One disadvantage of decision trees to keep in mind is that they do not work well with highly unbalanced classes. We do tend to place a lot more media behind certain media types and markets, which could limit the power of the model. Although decision trees are not the most competitive of supervised learning methods, they are highly interpretable and can approximate human decision-making, factors that are of primary importance in this use case!