Evolve Churn Model Debrief

Based on initial exploratory analysis of the data available to predict churn, a new churn prediction model would provide mediocre results. However, the results could still be valuable in identifying some property owners who are likely to churn so we can take action to prevent the churn. We would need to test the performance of the current model to the new model to determine if the new model should replace the old one. We could set up a champion-challenger test and if the challenger significantly outperforms the current model during a one-month time frame based on a two-proportions test, then we should replace the old production model with the new one.

Increasing model accuracy could be done by finding variables that are more predictive of churn. Having conversations with the data team and customer care team could be helpful in identifying what those variables are and if they exist in our data stores. Further work could be done to engineer features from the data available to improve model accuracy, although my gut says that such efforts would produce only a small benefit. Further work could be done to test other models and algorithms to see if model accuracy would be improved; however, some of the best, state-of-the-art models (such as gradient boosting machine, random forest and catboost) were already used in the exploratory analysis. If no further improvements can be made, the new model as it currently stands could be tested against the current model in production.

Assuming that the new model will be put into production, we ought to be aware of some issues that could arise from deploying a new model. We ought to compare the predicted results with actuals at least once every three months or so to see if performance is as expected. The forces that drive a property owner to churn could be in the making for some number of weeks or months, so we should keep in mind that we need to wait a sufficient amount of time after the model predicts that a property owner will churn to see if they actually do. The new model was trained on historical data that had certain trends embedded within it. If the economy or market shifts in some way and historical trends are disrupted, then the model may not perform as well since it won't understand the new trends. A model refresh would then be in order. For the first several weeks after deploying the model to production, the model results ought to be monitored frequently, perhaps twice a week, to ensure that the model results behave as expected and that the model does not have any bugs to fix.