Predicting Churn: Proposal

Ashton Reed

Churn is a concern for most companies and a huge issue for some of those. High churn rates can cost companies tens of thousands and can be an issue both internally (employees) and externally (customers). There are plenty of articles out there to inform companies of the danger churn presents to their business. However, I'm sure most companies following a recurring revenue model have already figured this out for themselves and recognized what churn can do to their monthly/quarterly/yearly expected revenue. There are several operational changes that can be made to reduce churn, and I am working on a LinkedIn post about it. However, if you read my post and determine that you have made all the adjustments possible to better accommodate your customers and still would like to learn more about predicting churn, please read on.

It's important to remember that there is no one-size-fits-all churn model. The generalized model for this project is built around a publicly available sample dataset, but it is also possible to build a model based on data from a CRM, such as Salesforce. Please keep in mind that in order to build and model and fit it to data, you must first have data, and while it does not necessarily have to be the cleanest, it does at least need to be fairly accurate. You cannot build a good model on bad data. No matter how impressive the model may seem with its pipeline and hyperparameter tuning, it will not be very accurate if it is trained on inaccurate data. For example, if you are not regularly updating the information in your CRM with a client's final payment date once they have churned, it will be difficult to add that customer to the "Churned" column when building the model for your data. A churn model will never be 100% accurate, but the more [pertinent and accurate] the data with which you can build and train a model, the better your chances are that you have more accurate predictions.

Despite how it may appear at first glance, churn analysis is not a simple classification problem but can be considered, more specifically, an issue of survival. Using publicly-available sample data, I will build a model that utilizes survival analysis to predict whether or not a customer will churn. By predicting churn, this model will help companies pinpoint what changes need to be made in order to retain customers. The deliverables for this project are a series of Jupyter notebooks containing the code for the model, two milestone reports, a final paper, and a slide deck.

Data source: https://www.kaggle.com/raumonsa11/churn-telco-europa