An Analysis of Customer Bank Churn

Machine Learning Capstone Project: DTSC 691

By: John (Jake) Behler

Background

- I looked at banking customer data to predict and analyze customer turnover. This can be important for incentivizing customers who are at risk of leaving to stay at the bank. Retaining customers is important for profit because it is costly to sign up new customers. When customers leave the bank (especially premium customers), it cost money to recover the loss and find new clients. High customer churn puts a big dent into revenue and profitability suffers because of it. I looked at prediction models and visualizations to better understand the data and give insightful information to the bank analyze which clients are likely to leave the bank. This will help the bank implement strategies to keep those customers at the bank.

Steps

- Gather data from Kaggle
- Data preparation, data cleaning, and data transformations
- Machine Learning models to make predictions on data in Jupyter (Python)
- Create visualizations in Tableau to analyze data for finding trends and patterns
- Create a website and PowerPoint to present insight

Insight from Tableau

- Complain is highly correlated with churn rate (99.51%)
- Germany has double the turnover rate compared to Spain and France
- Inactive members have higher churn rate than active members in both males and females
- Churn rate is higher in females (a difference in about 9%)
  - Males tend to stay slightly longer at the bank then females (0.1 years ~ a month)
- Customers who had 3 or 4 products have very high churn rate (82% and 100% respectively) compared to customers who only had 1 or 2 products (27.71% and 7.6%)
- Credit Score didn't have a big impact on customer churn rate
  - Except churn rate increased significantly for client who had credit score under 395
- 45-65 years old is where the highest turnover rate comes from
- Slight decrease in churn rate as satisfaction rates increase (besides France which is opposite)

Takeaway

- From out ML algorithm, Complain has very high predictability for Exiting the banks
  - If our complain variable is known to us before the customer leaves, this is critical and valuable information. This can be predictable to know that a customer will likely leave the bank. However, if this is only recorded at the time of exit, then this will not lead to impactful decisions.
- We have learned from out Tableau Visualizations about which types of groups have a higher churn rate which is important information
  - Complaining, being in Germany, being an inactive member, being a female, having 3 or 4 products, having a credit score under 395, being between the ages 45-65, and having a low satisfaction score all increase the rate of churn.

Challenges & Reflection

- From the beginning, I could have looked at engineering more features to possibly add to my model
  - I tried different methods such as binning variables, however, I only briefly looked at polynomials and interaction terms between variables
- Poor model performances with Tenure as my response variable
  - I had difficulty finding linear and non-linear relationships with this response variable and choosing the appropriate models
- Grid Search and parameters
  - I did do a grid search of my gradient boosting model, but I could have looked more into tuning the parameters and a grid search for random forest or other models
- Next time for my Tableau dashboard, I would expand the level of interactivity to include more complex filters or additional layers.