

PRESENTS

Churn Detection Modeling

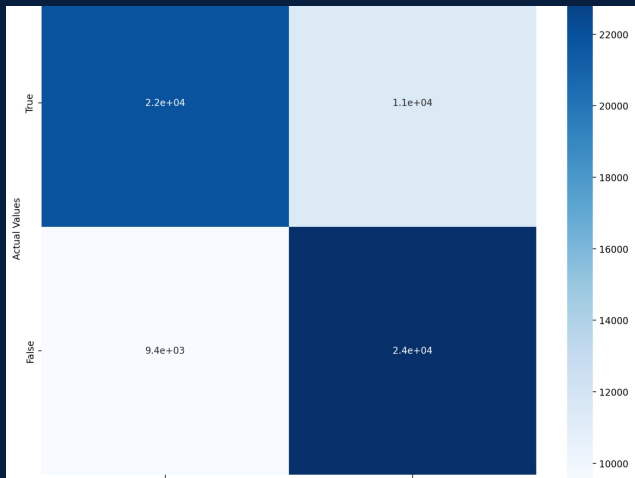
BY SYED HUSSAIN, MARCUS BEDEAU, REY JAIRUS MARASIGAN, KATERIN MARINAS, ERNESTO VILCHES MARTINEZ, CECILIA GERVASONI
MENTOR: LINA SOKOL, PRODUCT OWNER
INSTRUCTOR: DR. MASOUD SADJADI

Problems

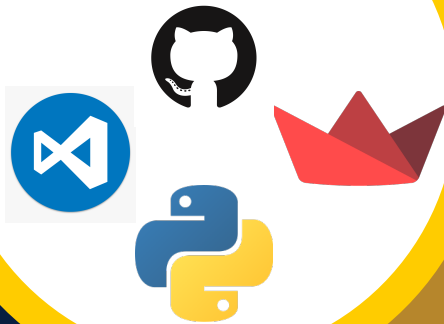
The company is interested in preventing voluntary customer departures as effectively as possible. For this, they are looking to implement a predictive churn model that estimates the likelihood of an existing client to leave so they can better target their rescue efforts. Our team is implementing a supervised machine learning model with a given data frame to best capture the relationship between the available features of the company and the target response. With our model, we will also present our findings in a visual platform.

Goals

To predict customer churn as accurately as possible, we cleaned our data by dropping some columns and joining columns from other csv's. Once the data was cleaned, we implemented stratified K fold cross validation and logistic regression to have an accurate model that identified churn factors. The model was progressively refined so that our graphs would depict and accurately identify possible churn factors. For better visualization, we utilized Streamlit which is an open framework that helped depict some of our results in a web application.



Implementation



Requirements

- * As a user, I want to have a homepage, so that I can view important information at a glance
- * As a developer, I want a single data frame, to make creating a machine learning model more efficient
- * As a user, I want to have multiple graphs that consume data from our model, in order to visualize customer churn.
- * As a developer, I want to have a correlation matrix of the data, so that we can assess relevant data.
- * As a developer, I want to be able to speed up the processing of our UI so that I will be able to see the data in a fast efficient manner
- * As a developer, I want to be able to improve the implementation of our bar graph so that data would be more easily readable
- * As a user, I want to be able to filter out certain columns of data that will reflect changes in the model

Summary

We took an assortment of data and ran it through our data cleaning python script. We then fed that data to a logistic regression model that would predict whether or not a customer would churn. The model was optimized to get the highest possible accuracy level but due to the insufficient amount of data we did not get as optimal levels as we had desired.

