

PROJECT

Predicting Customer Churn

Zahra Sameer & Ghadeer Alsandi

Abstract

Customer churn refers to the process of identifying customer/ clients who will terminate their relations with an organisation. It is very important aspect of an organisation as it helps to measure the growth of the company.

The purpose of this project is to build a model to predict if a given customer will churn or not churn using various classification algorithms and techniques.

Design

I will build a model to answer questions like (Looking at this customers given data, will he/she churn in month)

My initial baseline model was with Logistic regression with an F2 score of about 0.4.

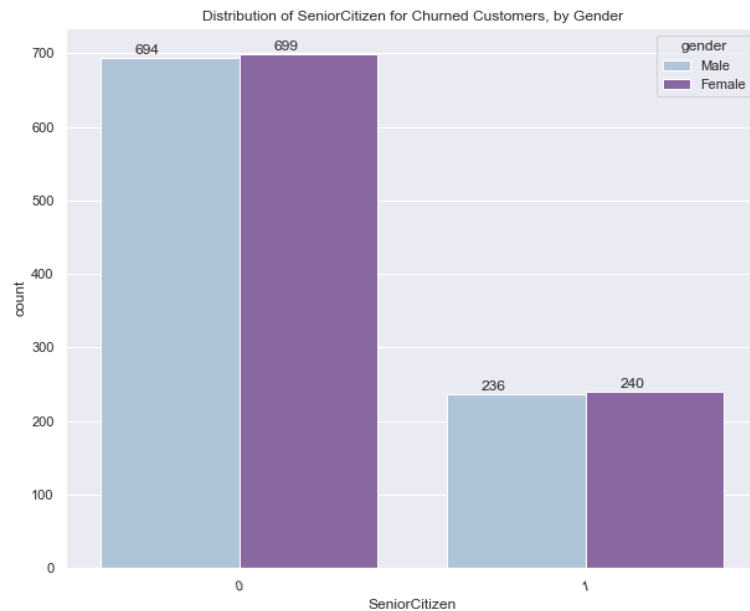
I built upon that by using various techniques like class imbalance techniques, cross-validation, and different models with hyperparameter tuning to achieve an optimal F2-score which we are about.

Data

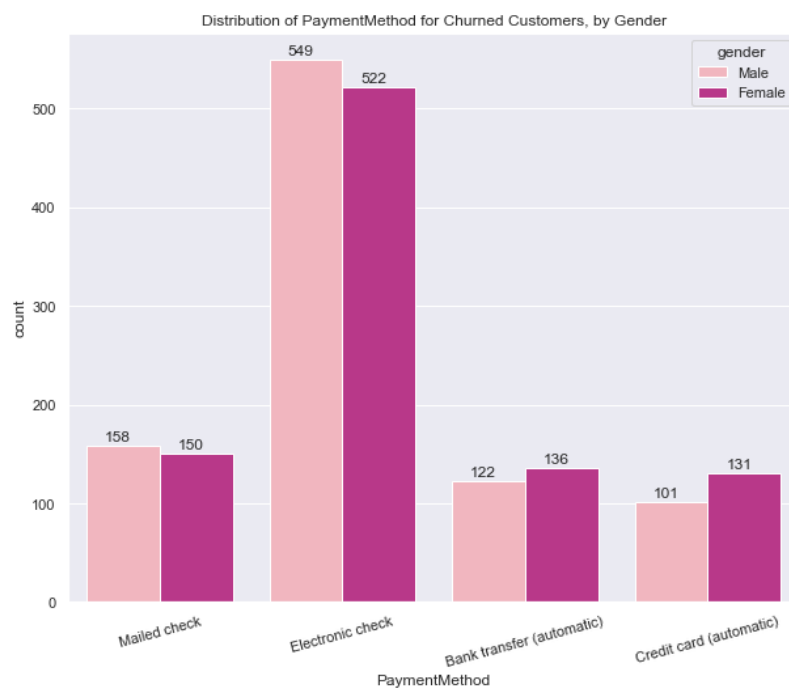
- The data for this project will be obtained from [IBM Sample Data Sets]
- Data will contain about 7,000 rows
- One individual row in the data signifies one unique observation of a customer's behavior (features)
- Customers who have left in the past– the column is called Churn.

Algorithm

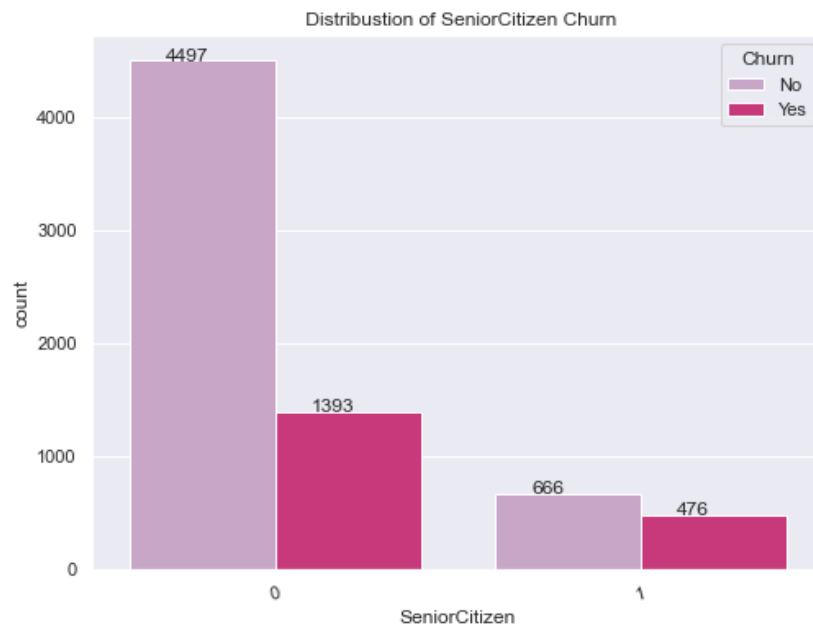
- We can also see from the barcharts below that, while performing a univariate analysis on gender gave little insights, infact almost even ratio.



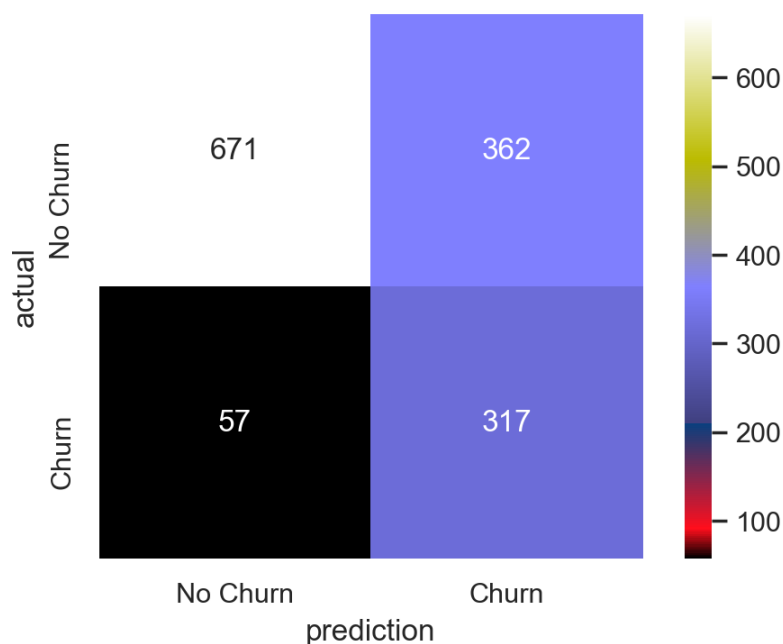
- Bivariate analysis reveals that males with electronic check payment type churns the most.



- In the figure below, our univariate analysis revealed the rate of churn among senior citizens is very high, compared to non senior citizens.

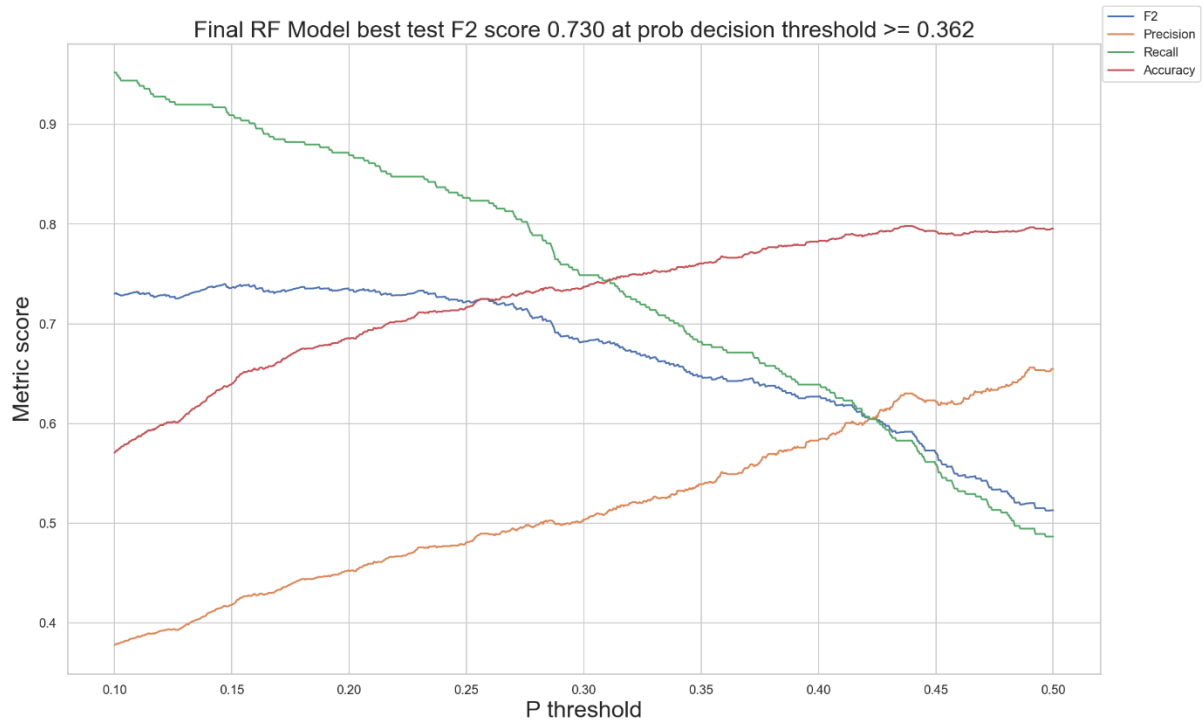


The main algorithm was a predictive classification models that was able to give as much accuracy score as possible. I ended up with an R2 score of .7 with a decision threshold at .22



- we capture as much turbulence as possible, so we'll make recall and f2 our priority. While we don't want to expect disruption when the customer wants it, as we don't want to waste resources on customers,

Here are some sharable models that have been tried including the selected random forest model with hyperparameter tuning.



Tools:

- Python
- pandas
- SQL
- Data Acquisition and storage and Modeling.