



Using customer behavior Data to improve customer retention

Final Project — DS-SF-44
Alfredo Rojas

San Francisco, May 2018

Using customer behavior Data to improve customer retention

What is in the Telco Customer Churn data set?

The data set used provides info to help predict behavior to retain customers.

The main purpose is to analyze all relevant customer data and develop focused customer retention programs.

Churn problem in telecom company. They need to understand who is leaving

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

Problem to solve:

Find out who are the type of customers that are leaving and create strategies that can enhance the experience

Hypothesis:

Are the customers that signed for a month contract churning more than the ones with 1 or 2 year contracts?

Using customer behavior Data to improve customer retention

Data Example:

https://community.watsonanalytics.com/wp-content/uploads/2015/03/WA_Fn-UseC_-Telco-Customer-Churn.csv

customer ID	gender	Senior Citizen	Partner	Dependents	tenure	Phone Service	Multiple Lines	Internet Service	Online Security	Online Backup	Device Protection	Tech Support	Streaming TV	Streaming Movies	Contract	Paperless Billing	Payment Method	Monthly Charges	Total Charges	Churn
6823-SIDFQ	Male	0	No	No	28	Yes	No	No	No internet service	One year	No	Credit card (automatic)	18.25	534.7	No					
9764-REAFF	Female	0	Yes	No	59	Yes	No	No	No internet service	Two year	No	Bank transfer (automatic)	18.4	1057.85	No					
0827-ITJPH	Male	0	No	No	36	Yes	No	No	No internet service	Two year	Yes	Credit card (automatic)	18.55	689	No					

Steps to follow:

1. Obtain Data
2. Examine Data
3. Check completeness
4. Exploratory Analysis
5. Define variables to use
6. Model creation and comparison
7. Conclusions

Using customer behavior Data to improve customer retention

In the dataset there were included 7032 rows of complete information

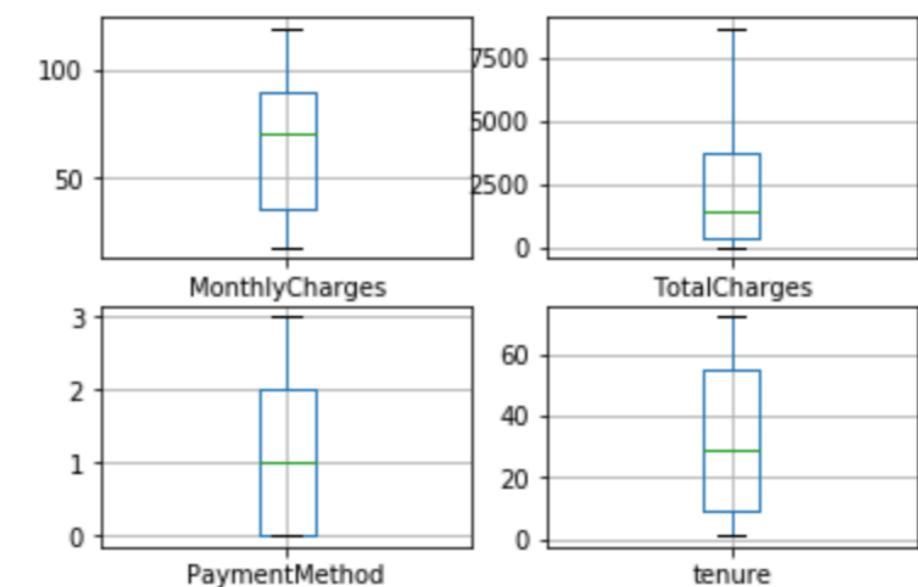
The variables were classified in:

- Time features
- Money features
- Phone features
- Internet features
- Demographic features

In order to do the exploratory analysis some transformations had to be done.

In this phase many different segments were created in order to understand the data better

As an example we could see the main time/money variables from the data set that already was giving information about the range of value this features helping to create segments around it

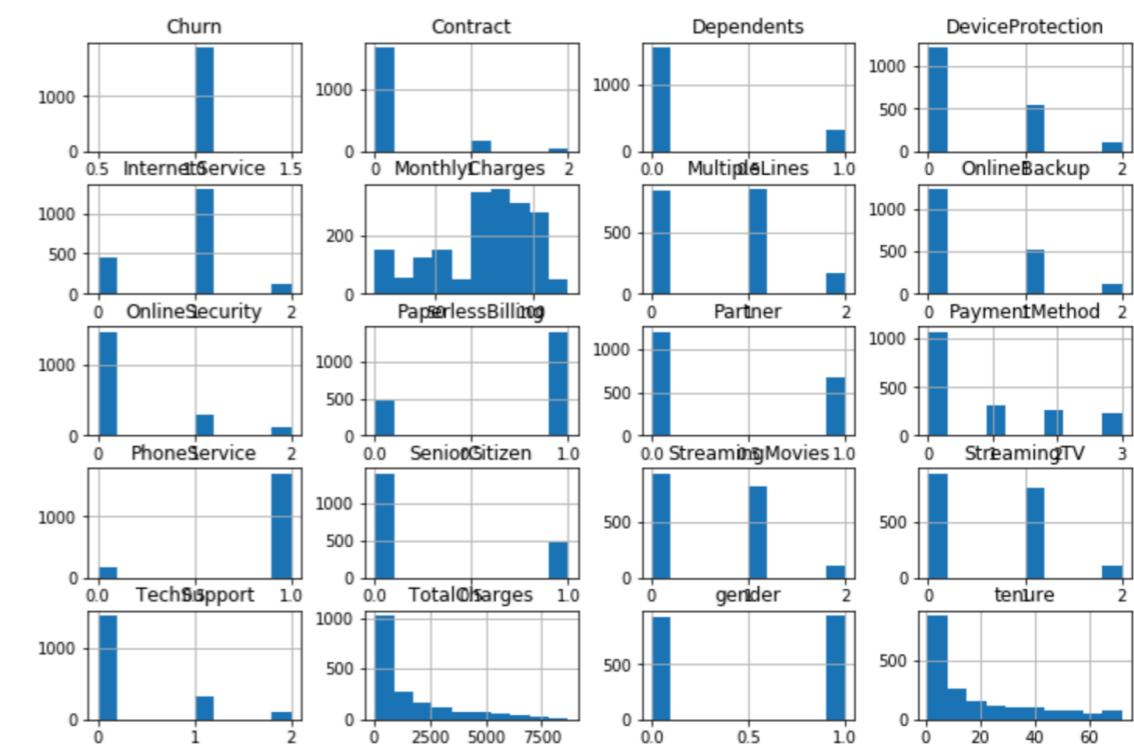


Box plot of time/money features

Using customer behavior Data to improve customer retention

This helped to create segments to visualize the data better and start driving conclusions:

- The majority of the users are in a month to month contract.
Median tenure: 30 months
- Big set of users that paying <30 dollars, the median amount for monthly charges is ~60.
- On one side the users that are paying the highest are Internet users and more towards the fiber optics side with a high tenure of the service and more equally distributed between the different types of contract.
- On the opposite side of the users whose totals are lower than the monthly median it can be seen that they are mainly users of DSL, no phone service and almost everyone in month to month services. On top of that the are churning almost 50/50 in the distribution.
- The conclusion of this is that one of the main pain points regarding churning can be found in the monthly contracts and the type of services they have.



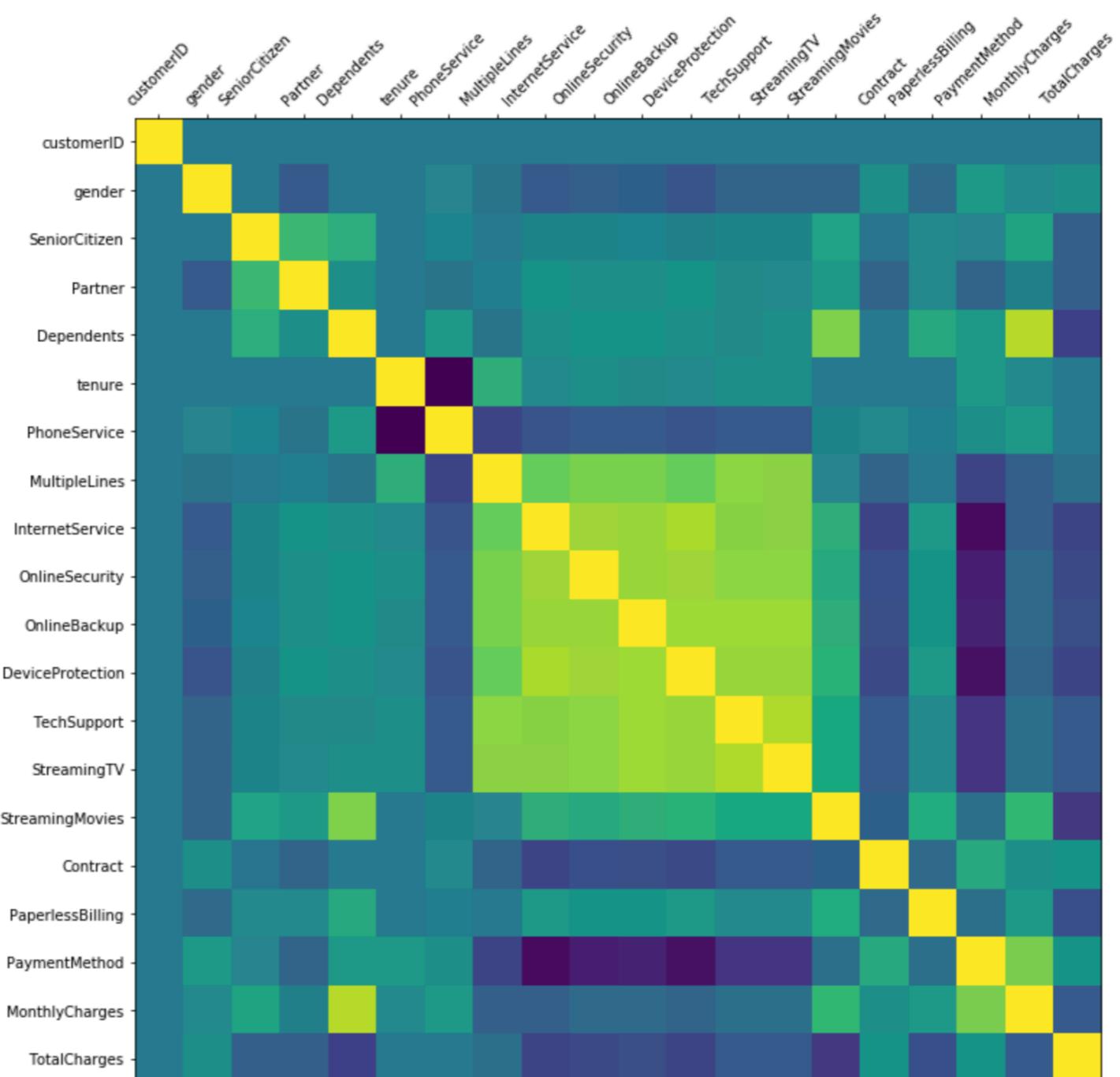
Histogram segment Churn=Yes

Using customer behavior Data to improve customer retention

One key step in the analysis was the correlation analysis

After doing this it was possible to understand that all internet features were related as same as the phone features and the total and monthly charges.

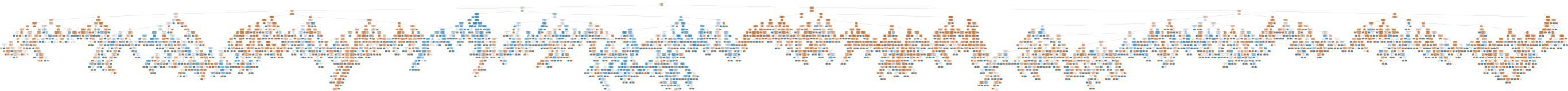
This made the analysis way easier in terms of model creation



Using customer behavior Data to improve customer retention

Model creation and Results

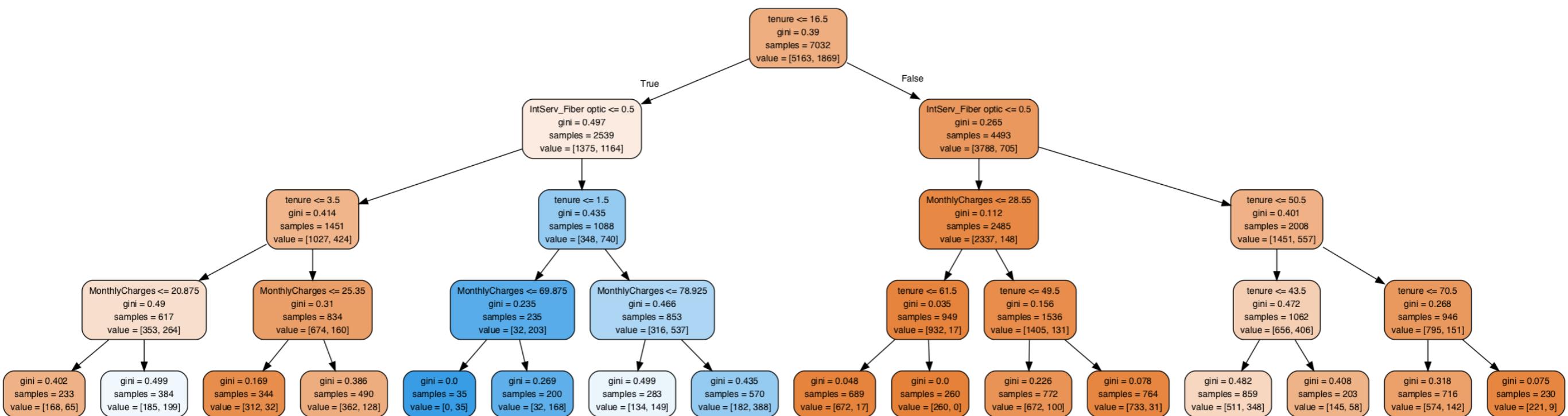
Model	ROC CV=5	Variables	Comments
Decision Tree	0.65	16	All features
Decision Tree	0.64	5	Top features only
Decision Tree max_depth = 4,	0.81	16	
Random Forest n=10	0.78	16	
Random Forest 10< n < 300	0.7827	5	n=161
Grid Search	0.78	5	n=81



Example of decision tree result with all features

Using customer behavior Data to improve customer retention

Best model result



Decision Tree with top features ROC=0.81

Using customer behavior Data to improve customer retention

Conclusions

- Different model were tested but the best result came from the adjusted decision tree.
- Even though that it was thought that less variables will increase the ROC it did not impact the results dramatically
- The random forest needs a lot of machine power to produce the same result as the adjusted decision tree
- One of the keys to model effectively this dataset was to test for multicollinearity, saving time, space and reducing the error.

Next Steps

- Creating solutions for the type of customers that the prediction might say that they will churn.