

Predicting Customer Turnover for Telecom company

BY: Carlos Marin

Outline

Background

Business Problem

Data

Methods

Results

Conclusions



Background

- Customer turnover affects every company. For this telecom company, the causes are first identified . Using machine learning identify models that provide the most accurate response for future predictions

Business Problem

- In a Perfect world, customer retention would be 100 percent. Unfortunately customer turnover is an everyday thing. First we identify causes of turnover. Then using machine learning algorithms, we can predict which would give us the most accurate predictions for future predictions

Data

- Using Telecom data set. Dataset has 3333 rows, and 17 columns.
- Contains data including
 - daytime calls, minutes and pricing.
 - night time calls, minutes and pricing.
 - International calls, minutes, pricing
 - Customer service calls
 - If customer had International plan
 - Customer turnover, if the customer left or still a current customer

Methods

- First evaluation using statsmodel logistic regression, to determine if variables in question have statistical significance to target variable
 - Non-Necessary variables were removed.
 - Developed Several Models to evaluate performance
- For Logistic Regression only
 - Plot the Area Under The Curve using ROC to evaluate performance of model
- For KNN, Decision Trees, and Random Forests Models, and Logistic Regression
 - Calculated accuracy using the confusion matrix
- For Decision Trees only, provide feature importance to see which features played a role in customer turnover/customer retention

Results

- Looking at Odds ratio data:
 - International Plan, has an odds ratio greater than one. They are 7 times more likely to leave.
 - Also customers who have made customer service calls are 55% more likely to leave.
 - Customers who make international calls are 8% less likely to leave, and customers who have a voicemail plan are 84% less likely to leave

Results

For Logistic Regression, the model had an accuracy score of 0.67, weighted low recall and high weighted precision

For KNN, the model had an accuracy score of 0.867

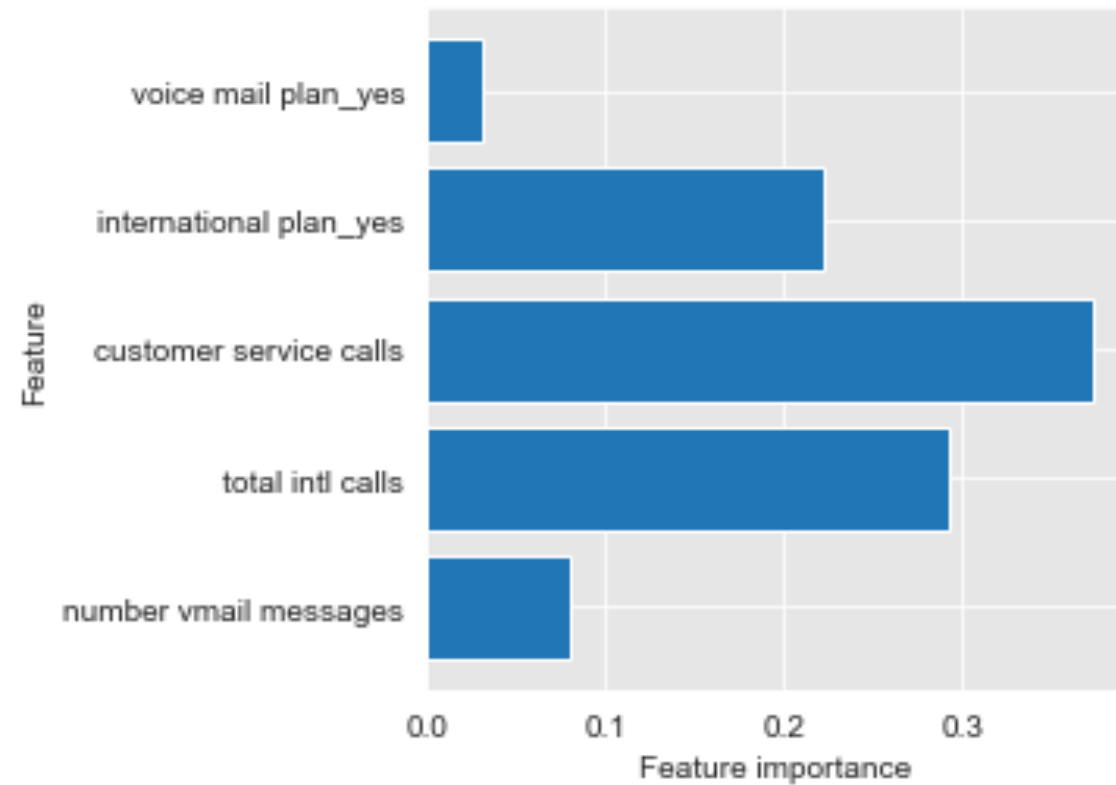
For Decision Trees, the model had an accuracy score of 0.868

For Random Forests, the model had an accuracy score of 0.867

KNN, Random Forests and Decision Trees all had high weighted precision and high weighted recall

Results

Decision Trees Feature Importance results



Plotting feature importance, using Decision Trees. customer services calls and then total international calls were top two important features, when determining customer turnover/customer retention

Summary of Findings / Conclusion

- Looking at Odds ratio, we see that International Plan, has an odds ratio greater than one. They are 7 times more likely to leave.
- Also customers who have made customer service calls are 55% more likely to leave.
- Plotting feature importance, using Decision Trees. customer services calls and then total international calls were top two important features for determining customer turnover / customer retention.
- Decision Trees has the best accuracy among the models tested when it comes to scoring for accuracy. KNN & Random Forests were a close second.



Thank you!