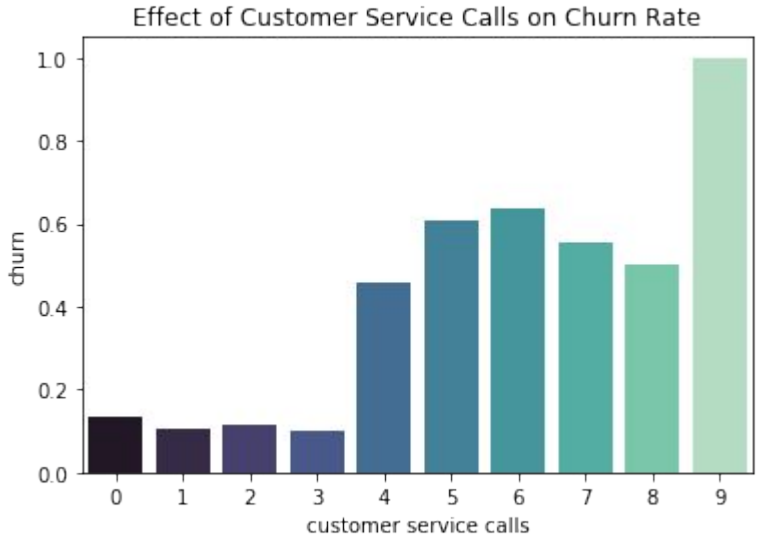
Predicting Churn in Telecom dataset

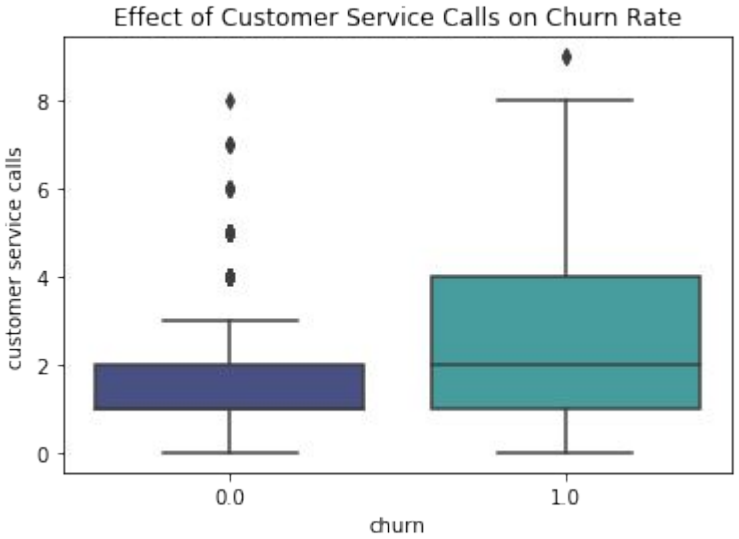
Jonah Flateman
6/2/21

Goal: Building the best classification model

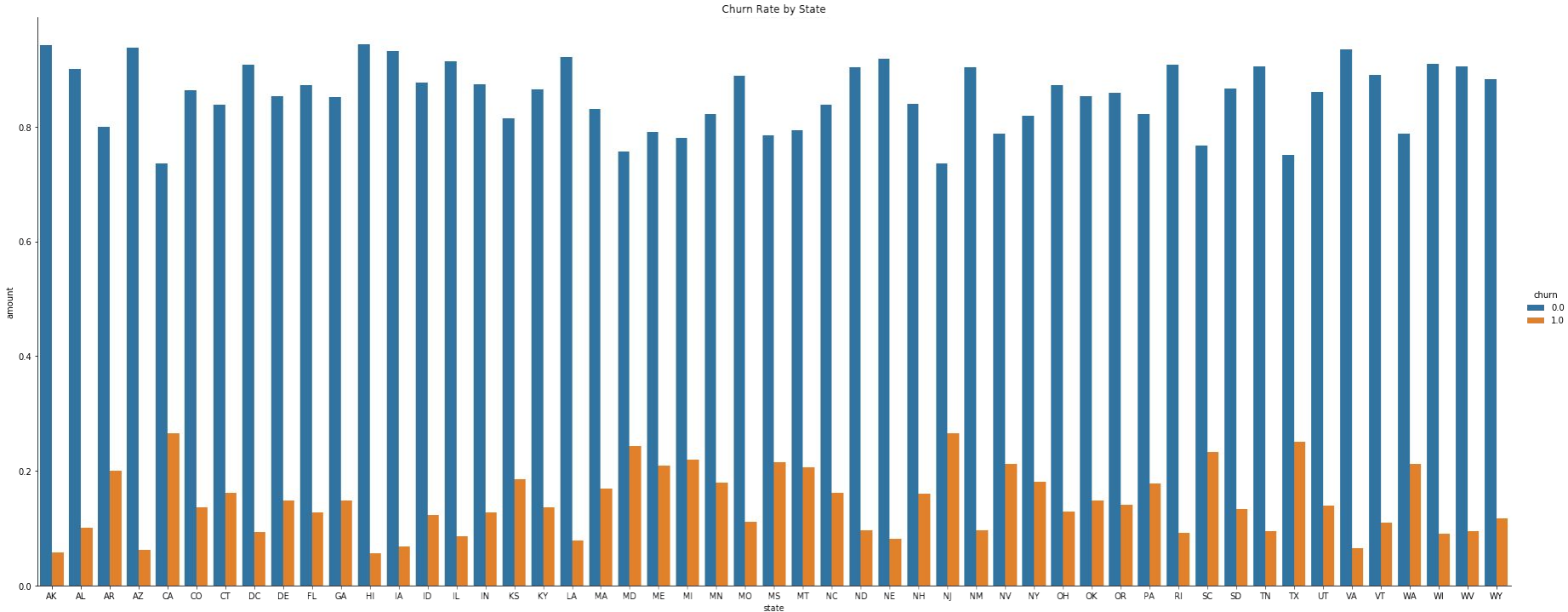
- Examine dataset and identify best predictors
- Start from baseline model and work through several iterations
- Visualize prediction results and feature importances

Exploratory Data Analysis: Visualizations





Exploratory Data Analysis: Visualizations

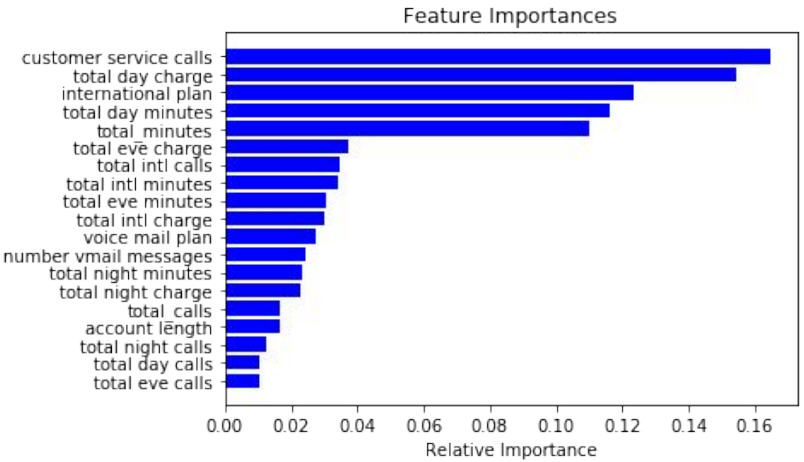


Modeling Overview

- Start with satisfactory model to work from
- Focus on recall and false negatives - why?
- Build two high quality models to work with and compare against each other

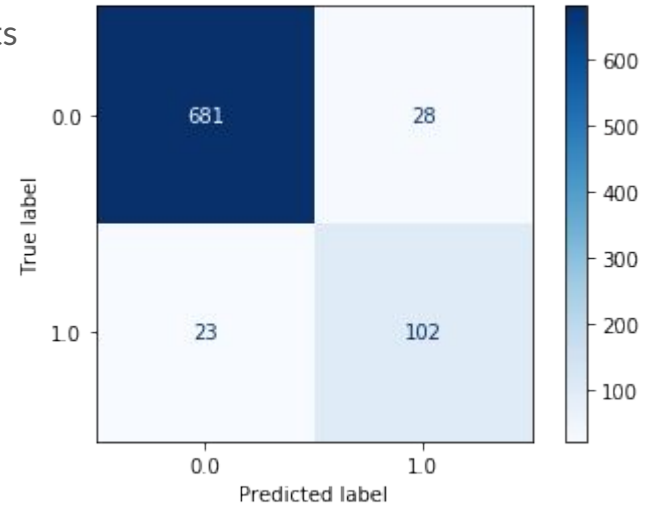
Model #1 - Random Forest Classifier

- Consistency between training and test data
- Relation to feature importance



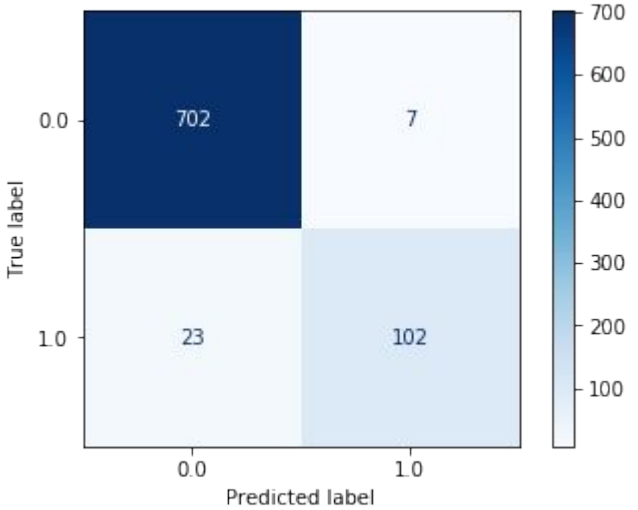
Model #1 - Random Forest Classifier

- Confusion matrix (for all classifiers) helps visualize results
- Keeping on eye on false negatives (lower left corner)

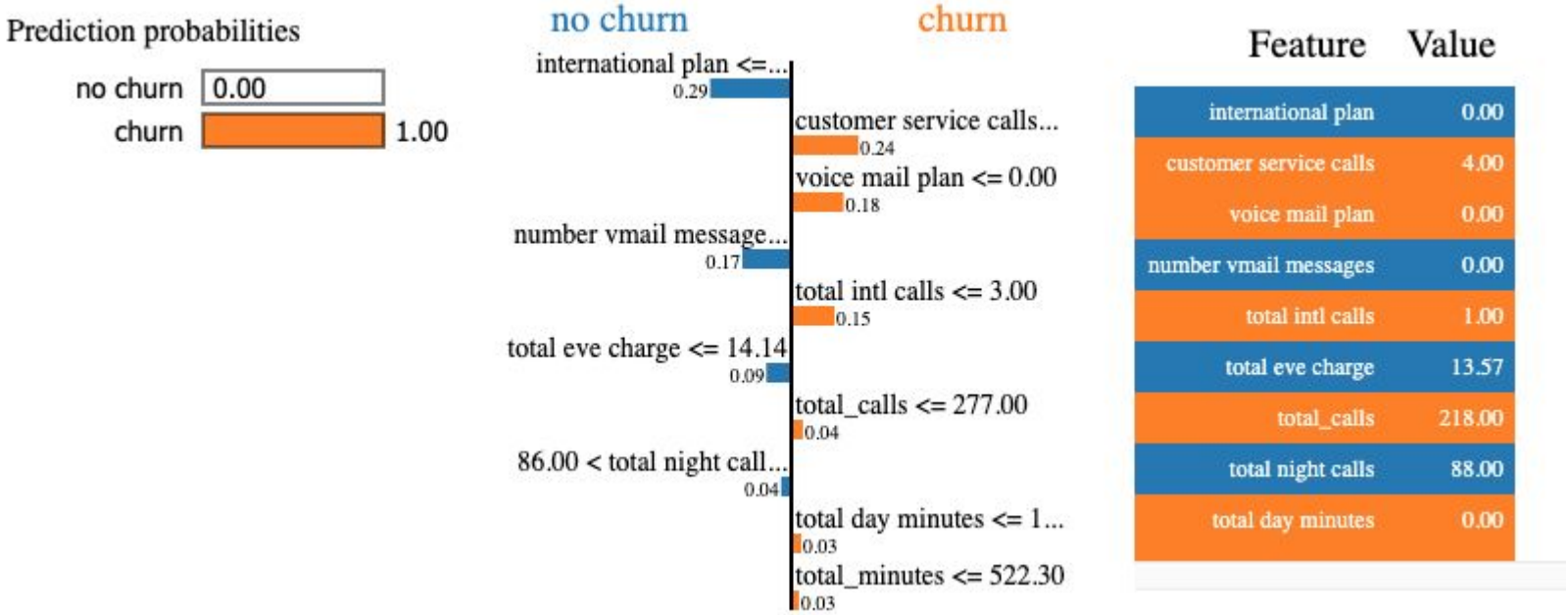


Model #2 - Gradient Boosting Classifier

- Slightly better trained model, fewer false positives
- Test results comparable with Random Forest Classifier
- Higher prediction accuracy



Model #2 - Gradient Boosting Classifier - LIME Visual



Future Work

- Reducing customer service calls
- International plans for international callers
- Adjust rates for day calls as needed

Thank You

jonahflateman@gmail.com