

Telecom-Churn-Prediction-Model

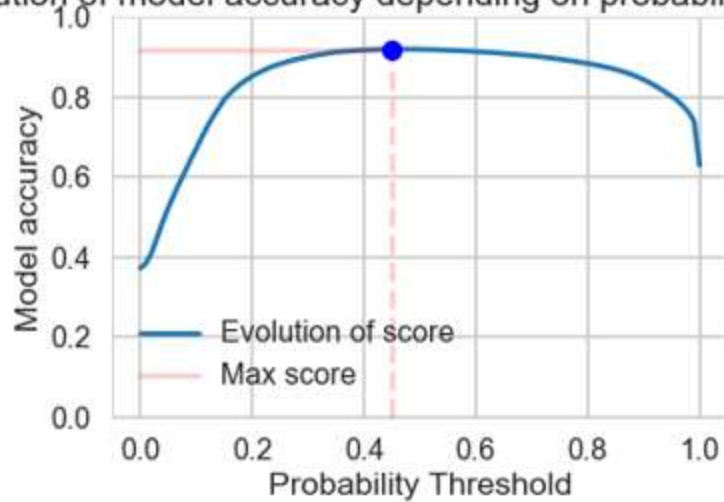
###Background

The goal of this project was to identify which customers would unsubscribe (churn) from a regional cell phone company. The telecom industry has been an area with a large amount of analytic research dedicated to customer retention, due to the large amounts of data generated and the saturation of the marketplace. Customer retention is five to six times cheaper than attracting new customers, and a one percent increase in retention can raise share prices by five percent.

Our data was sourced from a regional cell phone company located in the Southeastern United States, and contains customer information from 1996 through 1999. The overall churn rate for the customer base was 37%. Certain demographics had notably higher or lower churn rates. For instance, customers far from the southeastern region churned at very high rates (70%+).

- Algorithm
 - Logistic regression
 - Logistic regression + GridSearchCV
 - SVR
 - Random forest
 - Max_features : sqrt, log2, None
 - Comparison of (RT+LR , RF, RF+LR, GBT, GBT+LR)
 - Neural network (20 iterations)
 - SVM
- Results
 - **Logistic regression**
 - Accuracy on training data = 0.917623
 - Accuracy on test data = 0.918400

Evolution of model accuracy depending on probability threshold



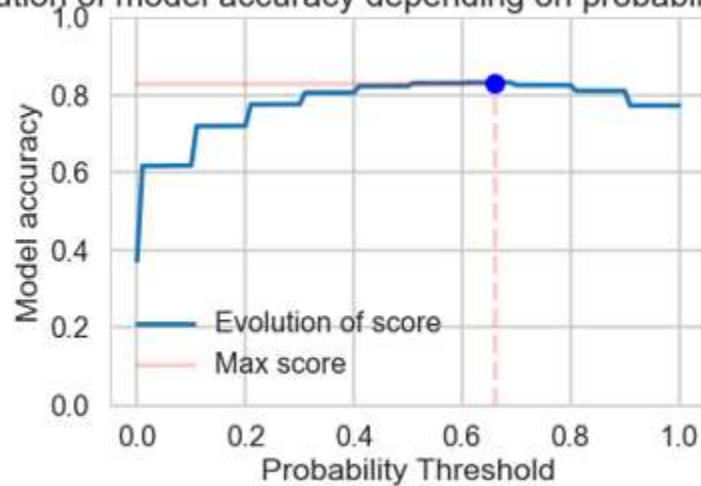
- Logistic regression + GridSearchCV

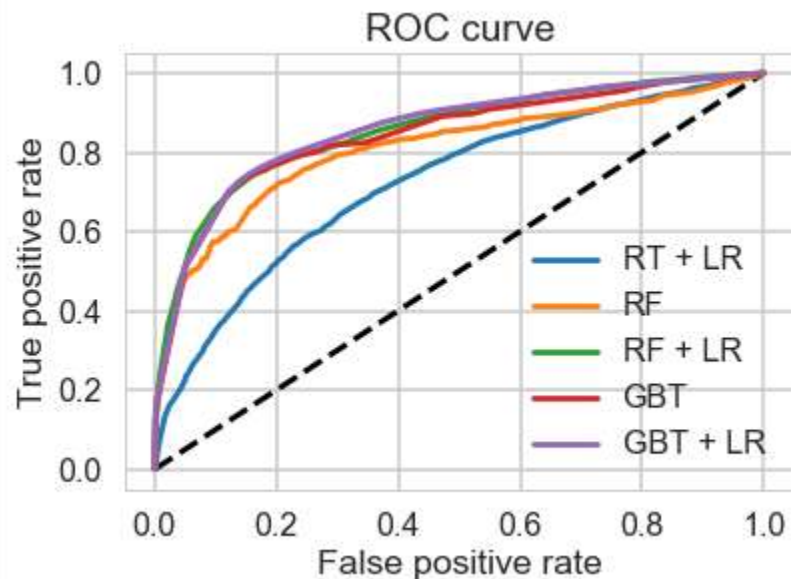
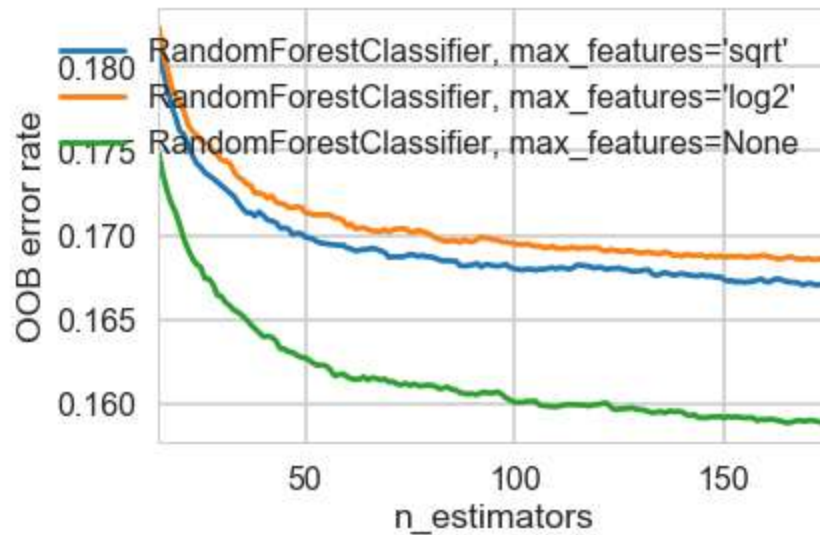
Accuracy: 0.9096750776345329

- **Random forest**

- Accuracy on training data = 0.980329
- Accuracy on test data = 0.829863
- Max accuracy occurs at probability threshold 0.66, with score of 0.83

Evolution of model accuracy depending on probability threshold



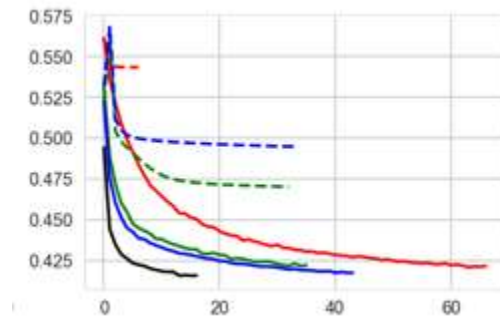


○ Neural network

- Training set score: 0.835857
- Test set score: 0.834312

-
- **training: constant learning-rate**
 - Training set score: 0.825486
 - Training set loss: 0.421454
 - training: constant with momentum Training set score: 0.825378
 - Training set loss: 0.422018

- training: constant with Nesterov's momentum Training set score: 0.825378
- Training set loss: 0.417291
- training: inv-scaling learning-rate Training set score: 0.740450
- Training set loss: 0.543167
- training: inv-scaling with momentum Training set score: 0.794553
- Training set loss: 0.470035
- training: inv-scaling with Nesterov's momentum Training set score: 0.776330
- Training set loss: 0.494604
- training: adam Training set score: 0.821191
- Training set loss: 0.415732



○ Non-linear SVM

