

Targeted Marketing In Telecom

USING MACHINE LEARNING TO IDENTIFY CUSTOMERS AT RISK OF CHURN

MICHAEL BOALS



Telecom Industry Churn

- ▶ Churn estimates range from 5% to 67% annually^{1,2}.
- ▶ 15 million customers change providers each year¹.
- ▶ Customer acquisition and retention spending is on par with infrastructure spending.

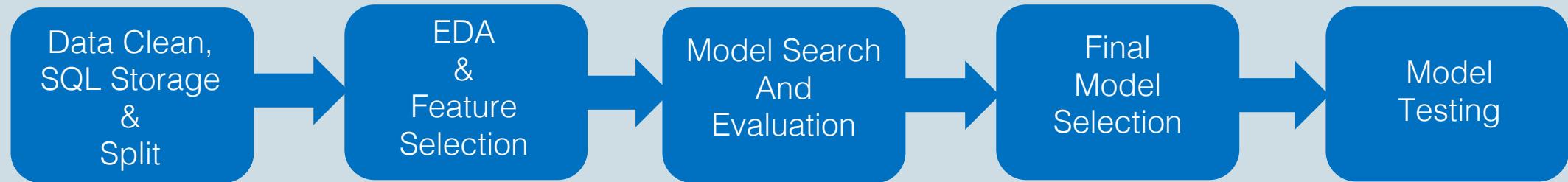


Project Goals:

- ▶ Use a small, representative dataset to build a computer model that identifies customers at risk of churn for use in a targeted marketing campaign.
- ▶ Prioritize correct identification of customers who might churn (Recall).



Methodology



Pandas



NumPy



Exploratory Data Analysis

- ▶ Dataset contained 7,000 entries for Telecom Customers
- ▶ Each entry had 20 features and churn outcome
- ▶ 27% of Customers Churned



EDA - Features

- ▶ 17 Features we used for model fit
- ▶ 15 Features are Yes/No binary values
- ▶ 88% of customers who churn are on month to month contracts.



Model Search

Models Tested

Logistic Regression

KNN Classifier

Random Forest Classifier

Support Vector Classifier

Naïve Bayes Classifier

Hybrid Model



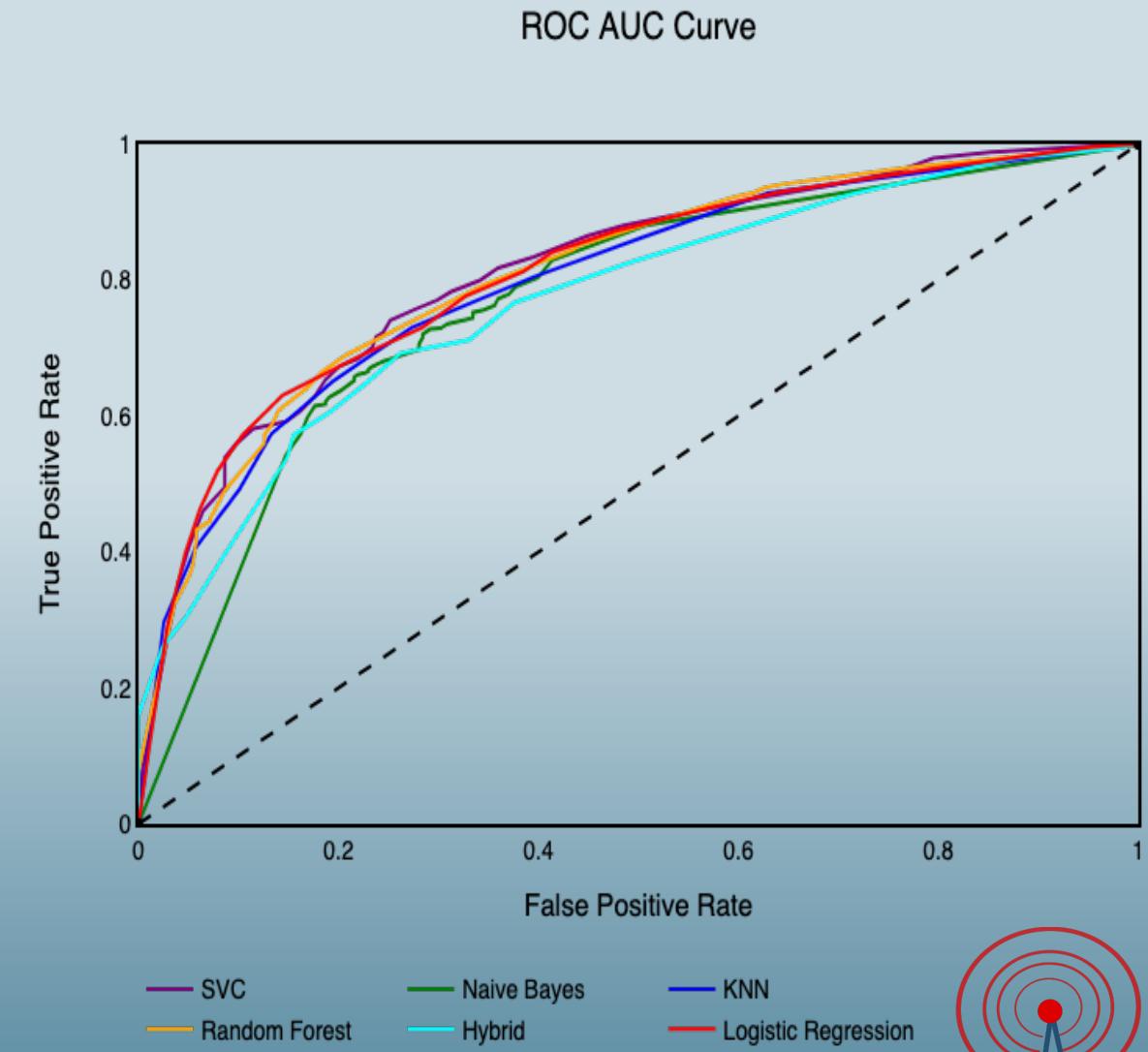
Model Search

Models Tested	Metrics	
	Recall	ROC AUC Score
Logistic Regression		
KNN Classifier		
Random Forest Classifier		
Support Vector Classifier		
Naïve Bayes Classifier		
Hybrid Model		



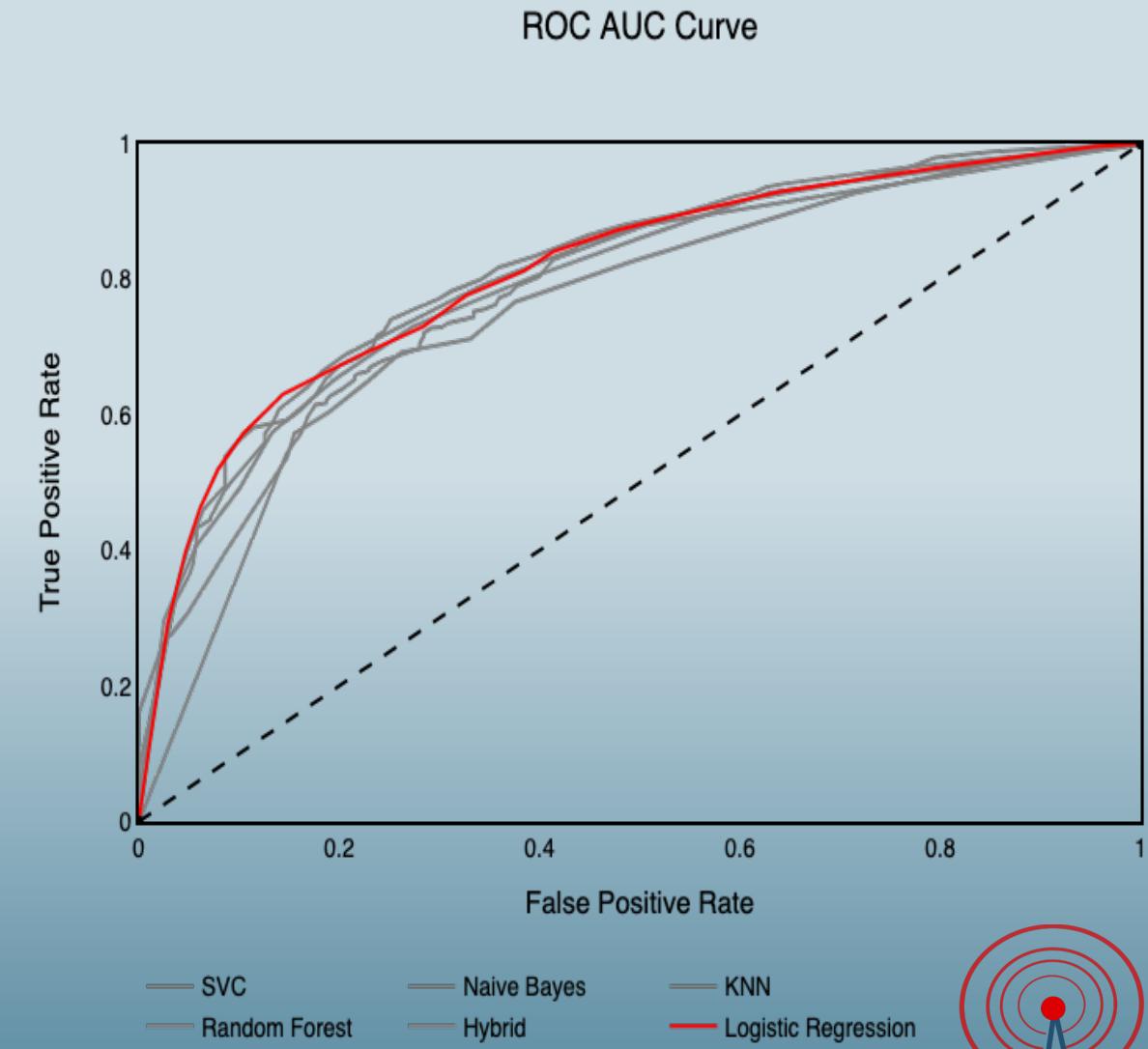
Model Search

Models Tested	Metrics	
	Recall	ROC AUC Score
Logistic Regression	0.76	0.83
KNN Classifier	0.75	0.81
Random Forest Classifier	0.86	0.82
Support Vector Classifier	0.78	0.82
Naïve Bayes Classifier	0.78	0.77
Hybrid Model	0.70	0.77



Model Search

Models Tested	Metrics	
	Recall	ROC AUC Score
Logistic Regression	0.76	0.83
KNN Classifier	0.75	0.81
Random Forest Classifier	0.86	0.82
Support Vector Classifier	0.78	0.82
Naïve Bayes Classifier	0.78	0.77
Hybrid Model	0.70	0.77



Final Testing Results

- ▶ Scores with Test Data
 - ▶ ROC-AUC Score: 0.86
 - ▶ Recall Score: 0.54

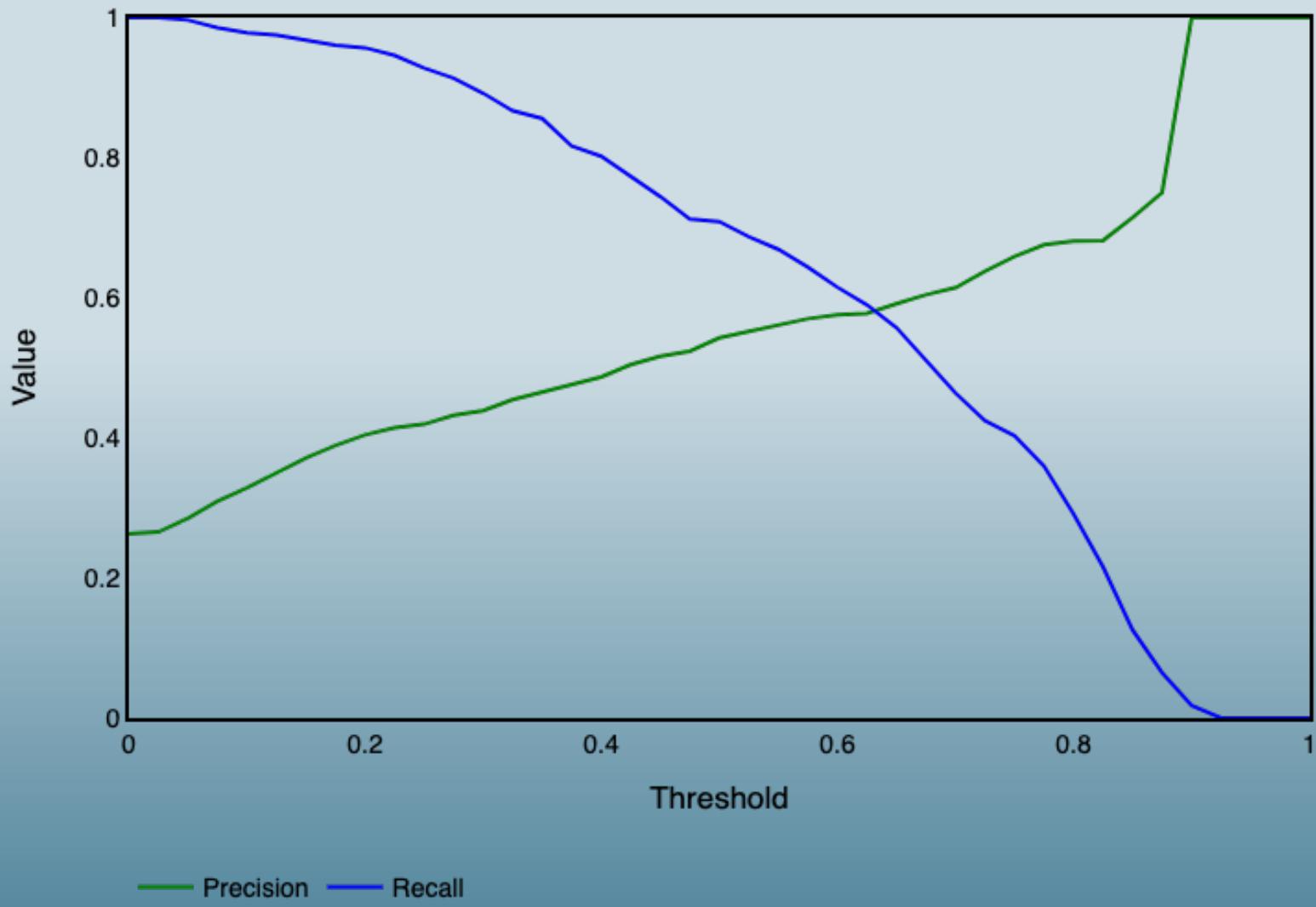


Conclusions and Next Steps

- ▶ Model Needs Improvement
 - ▶ Look for a broader feature set.
 - ▶ Consider ensembling models.
- ▶ Reoptimize model using true business impact.



Precision-Recall Plot For Logistic Regression Model



Questions?



References:

- ▶ 1. <http://www.dbmarketing.com/telecom/churnreduction.html>

- ▶ 2. <https://telecoms.com/opinion/churn-is-breaking-the-telecoms-market-heres-how-to-fix-it/>



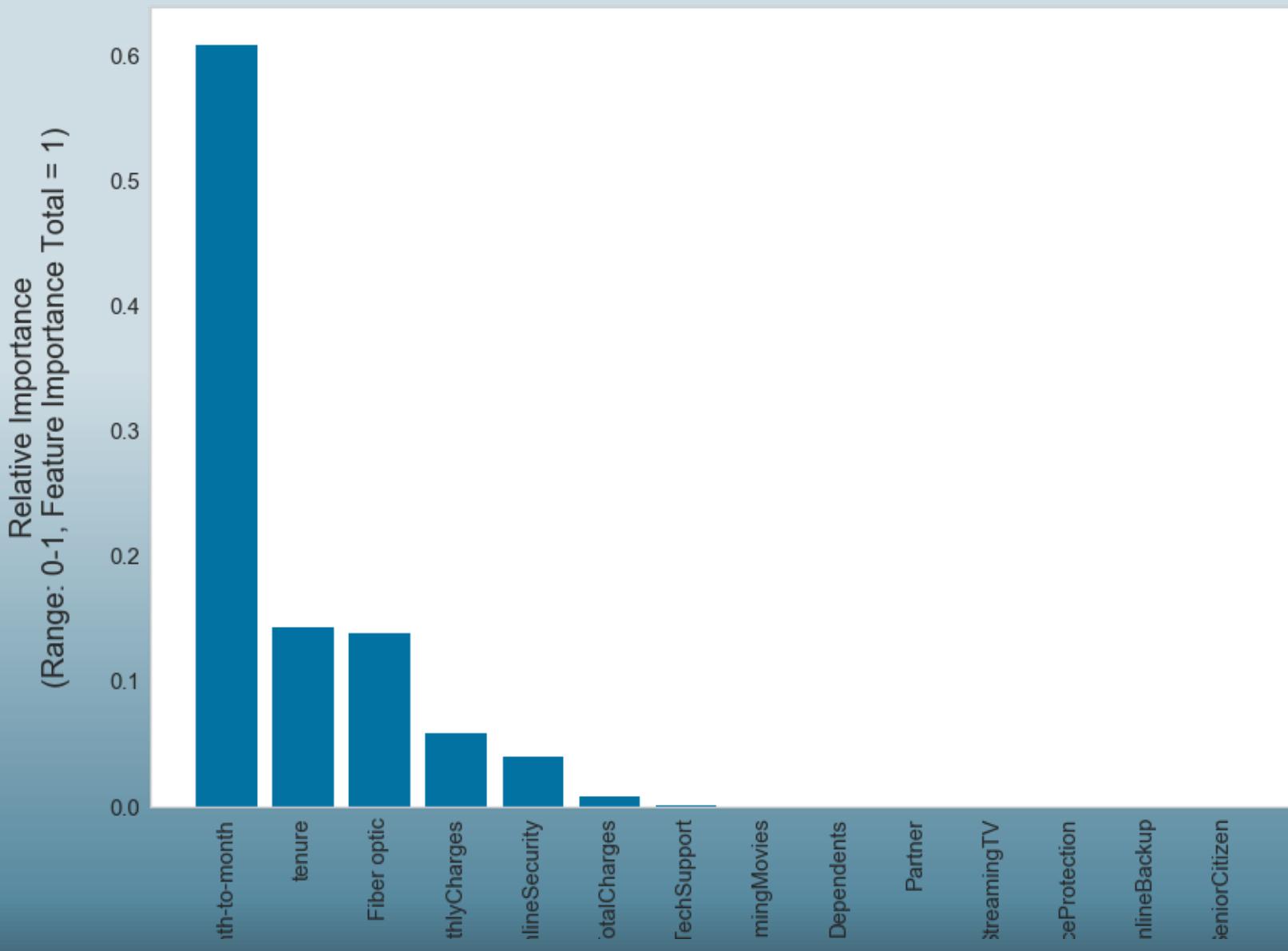
Appendix

- ▶ Graphs I didn't use in the presentation
- ▶ Financial Analysis
- ▶ Definitions of scoring metrics



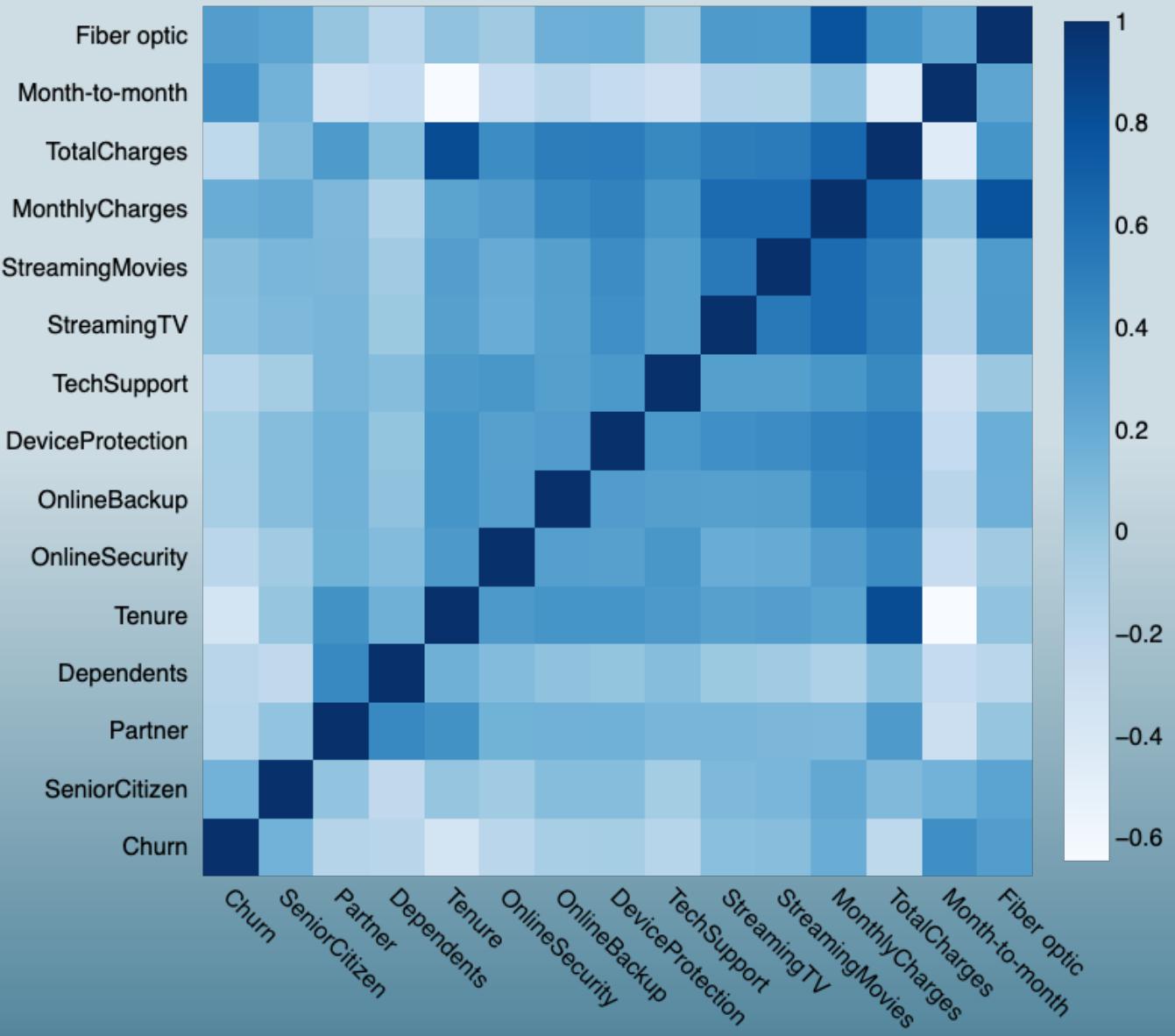
Feature importances

From Random Forest Classifier



Features

Correlation Heatmap



Complete Test Results

Models Tested	Metrics		
	F1	Recall	ROC AUC Score
Logistic Regression	0.60	0.76	0.83
KNN Classifier	0.53	0.75	0.81
Random Forest Classifier	0.61	0.86	0.82
Support Vector Classifier	0.63	0.78	0.82
Naïve Bayes Classifier	0.59	0.78	0.77
Hybrid Model	0.53	0.70	0.77



Model Scoring

$$\text{True Positive Rate (recall)} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{False Positive Rate} = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}}$$

$$\text{Positive Predictive Value (precision)} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

$$\text{F1 Score} = \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$



Simulated Business Impact

Obviously doing nothing will mean significant loss of revenue for the company. Conversely, offering an incentive to all customers would result in unnecessary expenses for content customers. We need to find a happy middle ground.

At an average monthly charge of \$64.76 regular churn translates to an annual loss of \$121,036.

Building a financial model for an incentive plan might include a per customer incentive cost and an acceptance rate. The cost calculations are then:

$$\text{Churn Reduction} = \text{Base Churn Rate} * \text{Recall} * \text{Acceptance Rate}$$

$$\begin{aligned}\text{New Churn} &= \text{Base Churn} - \text{Churn Reduction} \\ &= \text{Base Churn} (1 - \text{Recall} * \text{Acceptance})\end{aligned}$$

$$\text{Incentive Cost} = (\text{tp} + \text{fp}) / (\text{tp} + \text{fp} + \text{fn} + \text{tn}) * \text{Acceptance Rate} * \text{Incentive Cost}$$



Plugging in promotion cost of 4% of monthly charges and an acceptance rate of 50 percent we get the following cost v model threshold graph.

Simulated Promotion Cost/Savings

