The Great Hackathon of January 2019: Predicting Churn

Emily J. Cain
&
Andrew Wester

Preprocessed the Data

- Hunted down the 11 missing values (in TotalCharges)
- Converted to boolean values when possible
- One-hot encoded remaining categorical variables
- Used StandardScaler

Optimized Models using GridSearch for Recall Score

Better to wrongly predict that some customers will churn, and end up staying, rather than miss a customer who will churn that we predicted to stay.

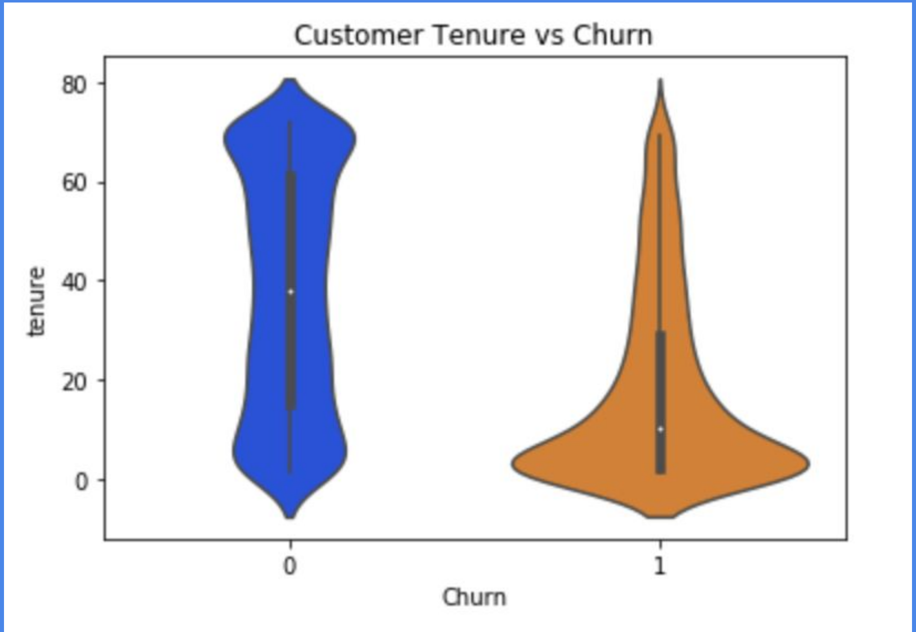
Recall Scores:

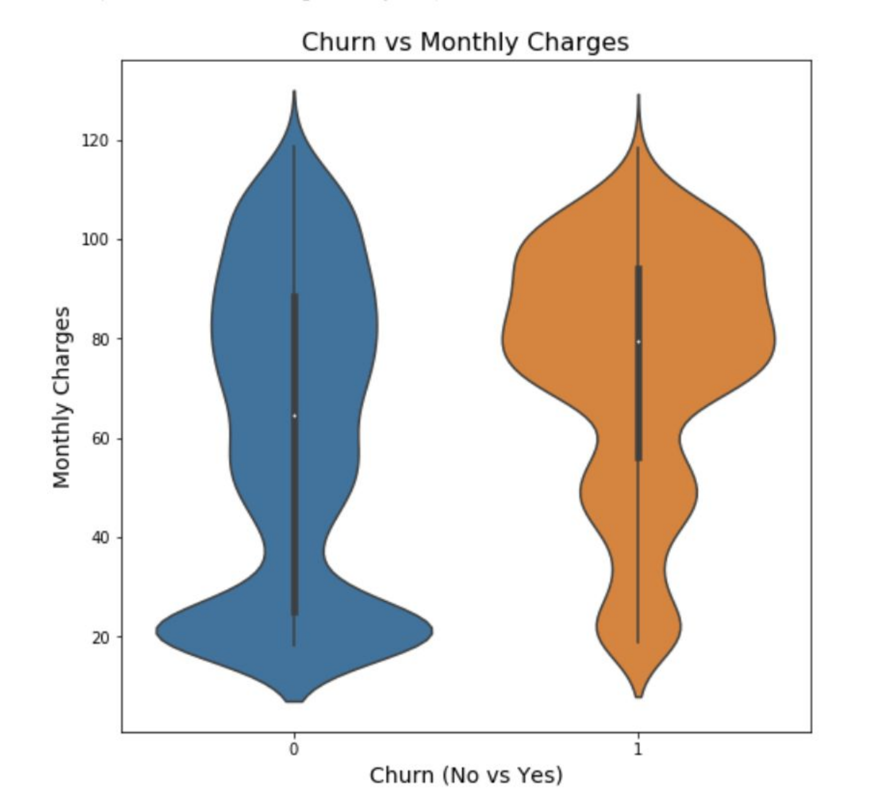
- Decision Tree: 91% Recall (Accuracy 67%)
- Random Forest: 77% (Accuracy 64%)
- Gradient Boost: 50% (Accuracy 80%)
- AdaBoost: 49% (Accuracy 78%)

IF YOU WERE DIAGNOSED AS NOT HAVING MYSTERY DISEASE, BUT ACTUALLY HAD MYSTERY DISEASE, THAT WOULD BE BAD → WE WANT HIGH RECALL

IF YOU WERE DIAGNOSED AS HAVING MYSTERY DISEASE, BUT ACTUALLY DIDN'T HAVE MYSTERY DISEASE, THAT WOULD BE BETTER THAN THE ALTERNATIVE → WE ARE COMFORTABLE WITH LOWER ACCURACY

Short Term Tenure & High Monthly Charges = Highly Likely to Churn!



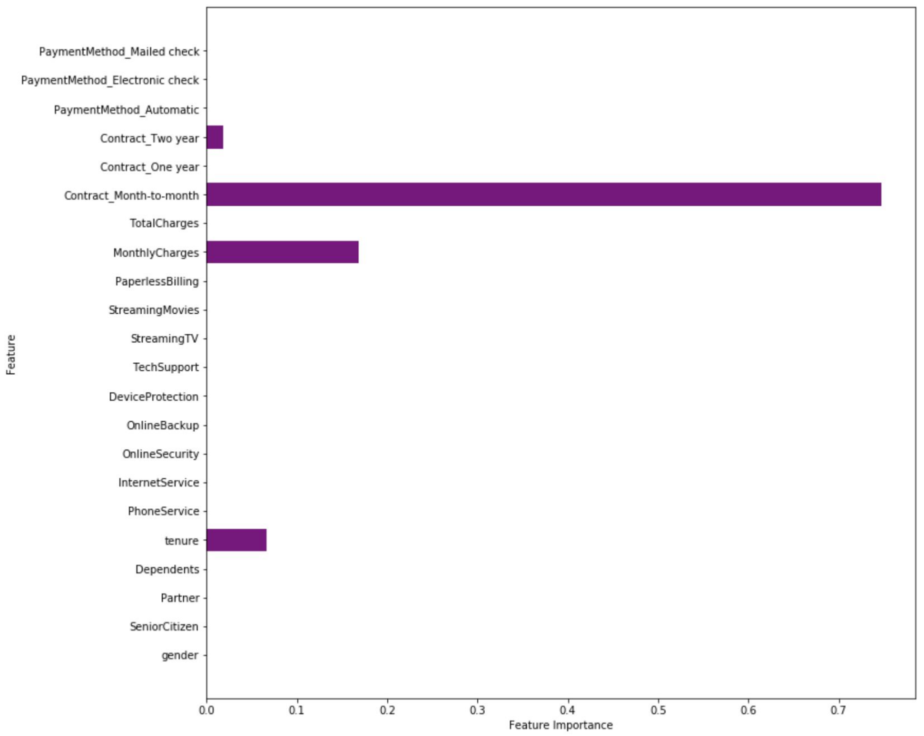


Most Important Features of Decision Tree

<u>Top Features:</u>

- Monthly Contract
- Monthly Charges
- Tenure

```
In [77]: plot_feature_importances(optimized_dt)
```



Targeting Customers Who May Churn

Our model shows:

- Customers most likely to Churn:
  - Have shorter tenure
  - Higher monthly payments
  - Month-to-month customers

RECOMMENDATIONS:

- Make longer contracts more attractive than monthly contracts
- Lower monthly payments to entice more customers
- Reward loyal customers