K Nearest

Eoin Flynn

18 March 2018

Bond University
Data Science

Contents

Introduction

In this report we will build upon the information gathered from our decision tree and logistic regression models to build a K-Nearest Neighbour model that can be used to predict whether a customer will churn. We will complete a series of tests to produce the best possible model using the customer dataset provided to us.

Model

After developing a series of K-nearest neighbour models we were able to create a baseline model and optimise it to best fit its intended use case. As you can see from the ROC plot below, the model has an overall accuracy of 83.11%, but more importantly the accuracy of the model's yes predictions are 71.88%. The model did produce 288 false positives but a misclassification that small should be of no real convern considering how accurate the model is at predicting yes all other times.

Model ROC
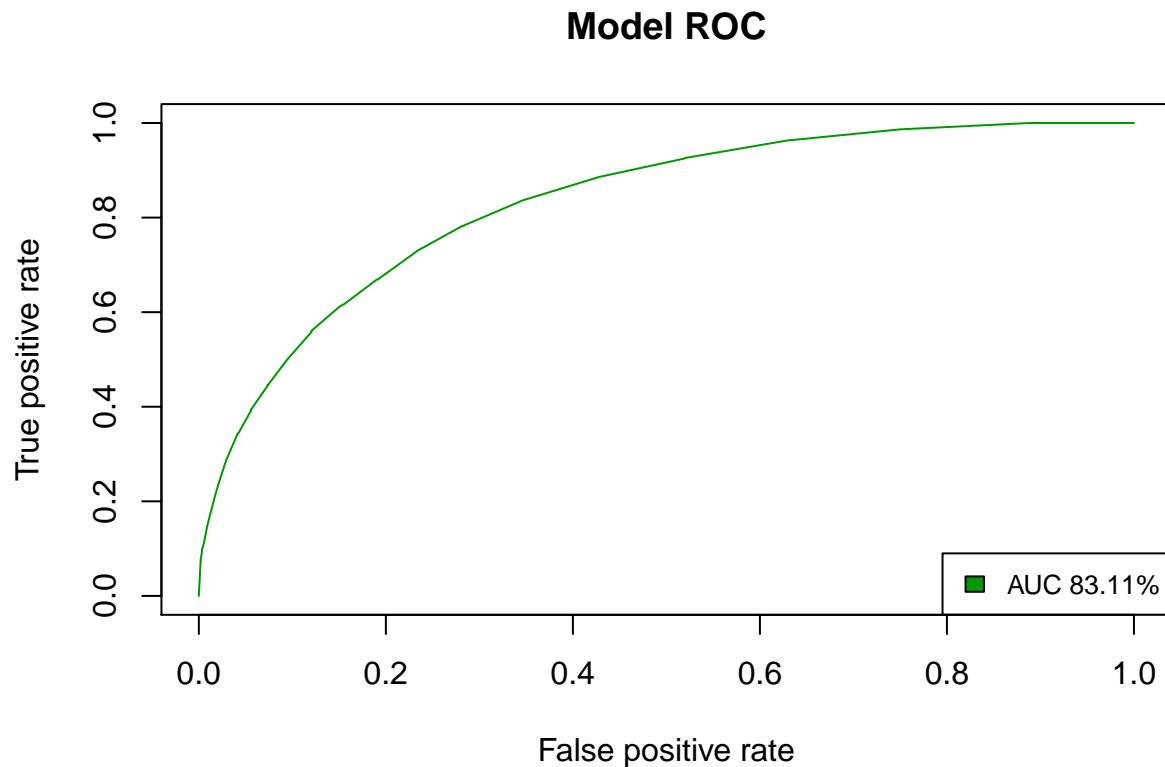
If we create a dummy customer we can test to see whether they will churn.

```
##   NegativeProb PositiveProb classification
## 1    0.3461538    0.6538462            Yes
```

The model predicts that a male who has a one year tenure, streams TV and has a fibre optic internet service will churn, and based on the results from our testing we can say that with a 71.88% chance of being correct. From an output such as this we would make recommendations to upgrade your fiber network or invest further into your tv streaming service.

Conclusion

Given the model's accuracy when it predicts that a customer will churn, we can advise that this model should be used in everyday business operations. There may be parameters outside of those provided in the dataset which would improve the accuracy of the model but that can be visited at a later time. For now this model will work for its intended business use.