



In the Game of Thrones you either win or you die...



# Supervised Learning

Supervised learning is a data mining task in which a machine is trained based on a number of labeled observations. By logging the outcome of every possible scenario, supervised learning allows a machine to make prediction based on precedent.

I will use supervised learning to train my model to make predictions on whether a character in the popular TV show, Game of Thrones, survives



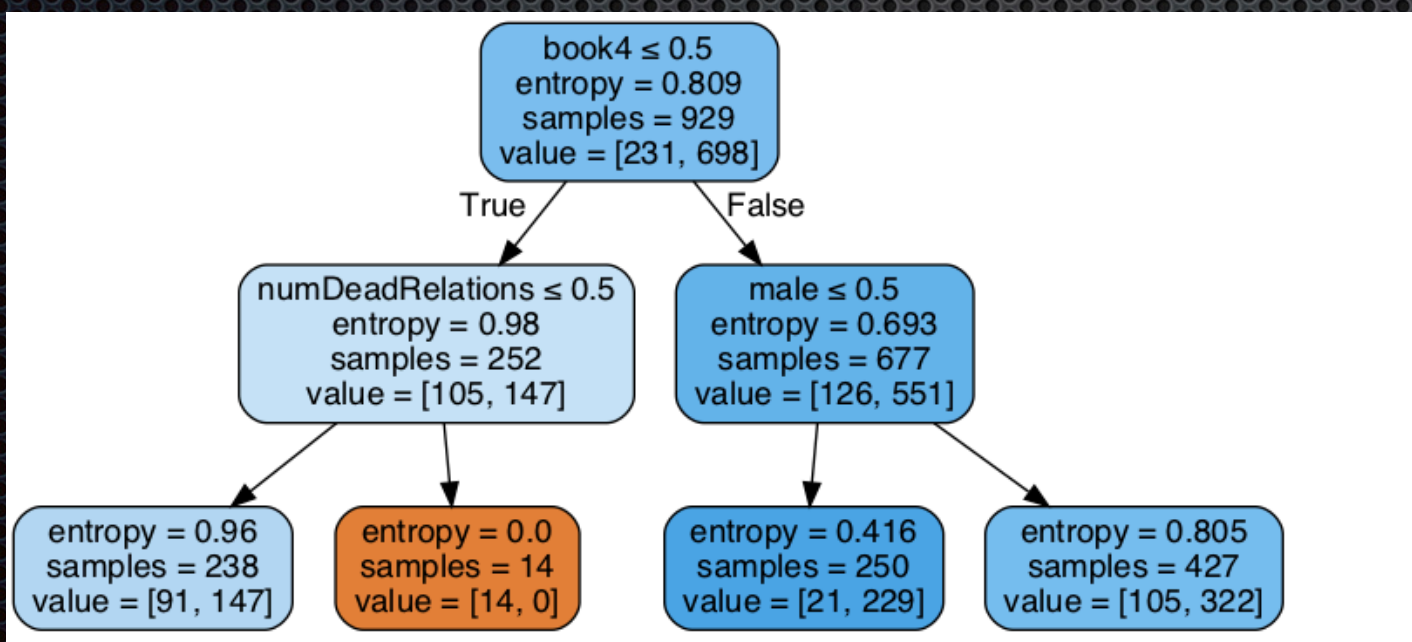
# The Predictors

- Gender: Male or Not
- House : Which house they were born in
- Book: Which book do they make an appearance in
- Married: Are they married?
- Nobility: Are they noble?
- Number of Dead relatives
- Are they popular? How popular?



# Decision Tree

- The decision tree model is a flow chart that uses the observed behavior of different features to make predictions on our target.
- This decision tree was able to predict with 80% Accuracy and 89% precision whether a character will survive or not
- To the left we see the decision tree made by our model and to the right a use-case example of this tree



Character: Ned Stark  
Not in Book 4?: True

Character: Ned Stark  
# Dead Relatives < 0.5?:  
False

Character: Ned Stark  
Alive?: False



# Random Forest

- ✦ This model makes use of several decision trees to make a prediction on a target
- ✦ The random forest model that I created is able to predict with 84% accuracy and 90% precision on whether a character dies or not



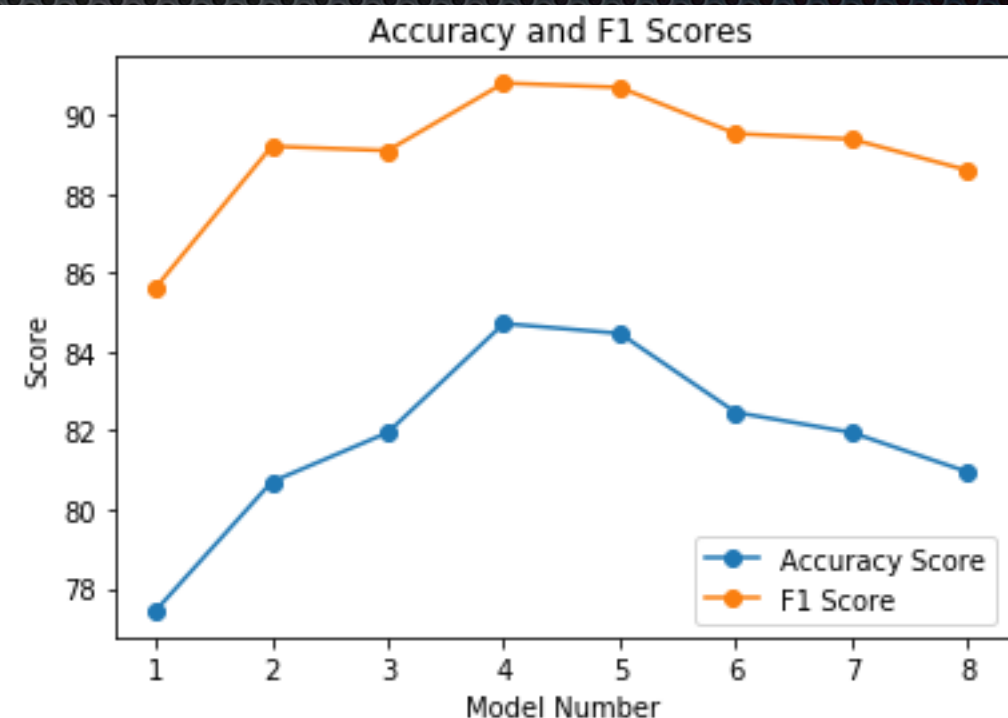
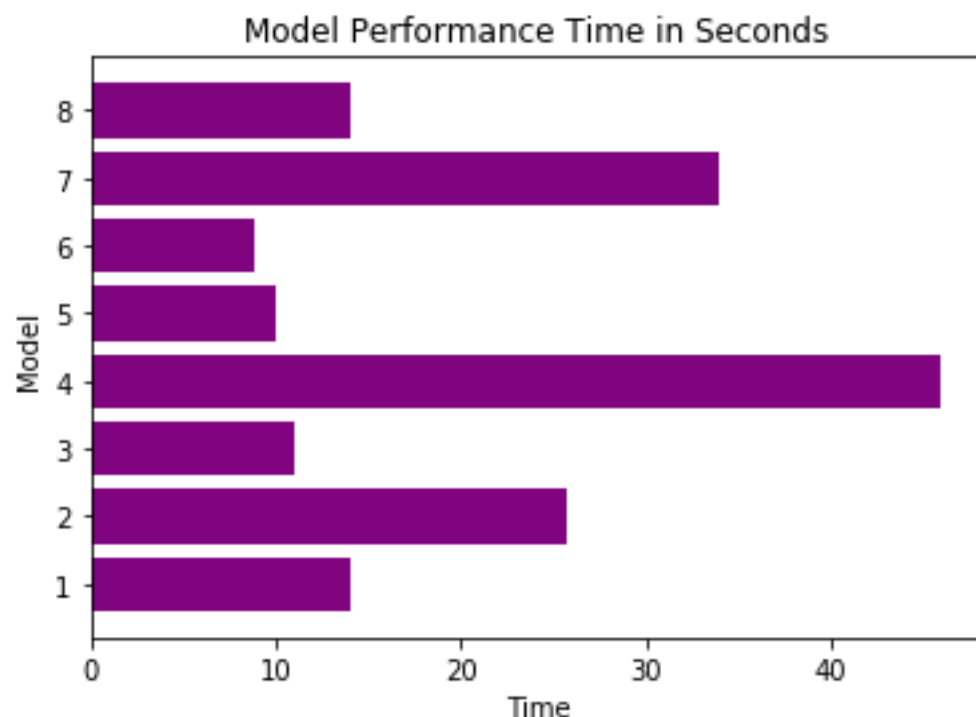
# XGB Boost

- ✦ The model I ultimately chose however, is the XGB Boost classification model.
- ✦ XGB is a high performing machine learning method which implements gradient boosting to make accurate predictions
- ✦ My XGB model is able to predict with 80% accuracy and 88% precision on whether a character survives or not. It was able to do so using 11 fewer features than any other model



# Choosing a Model

- ✦ In choosing a model, I examined which model yielded the highest accuracy and precision scores in the least amount of time.
- ✦ Model 5 & 6 were neck and neck on all three measures. However, Model five used 5 fewer features than Model 6 and was ultimately chosen as the most efficient model. Model 5 was a Random Forest and Model 6 implement XGB Boost





# Conclusion

- Decision classification uses a flow-chart observing all features provided to make predictions on a target
- Random forest classification uses several different decision trees to weigh in on which features hold more weight and uses its observations to make predictions on a target. This model reported the highest scores for accuracies and precision and made its prediction at the second fastest rate.
- XGB classification proves to be the model that can make highly precise and accurate predictions while using the least amount of features