

Utilizing Machine Learning with NFL Play-By-Play Data

Acusio Bivona





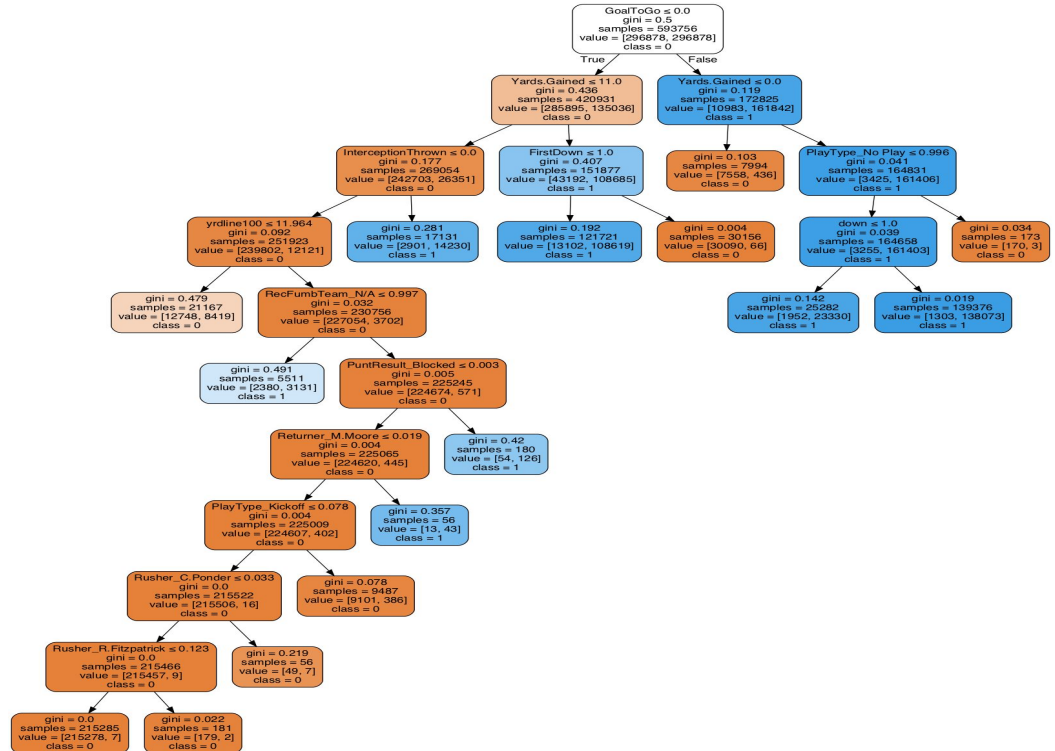
The Data and the Objectives

- NFL play data from 2009-2017
- Dataset:
 - 7,068 predictors
 - 407,688 entries of data
 - Total: 2.88 billion pieces of data
- Two objectives:
 - Classify a play as being a touchdown or not a touchdown
 - Determine most important predictors



Training & Testing = Machine Learning

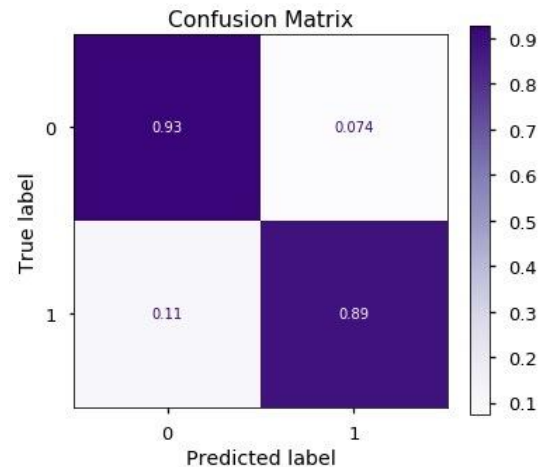
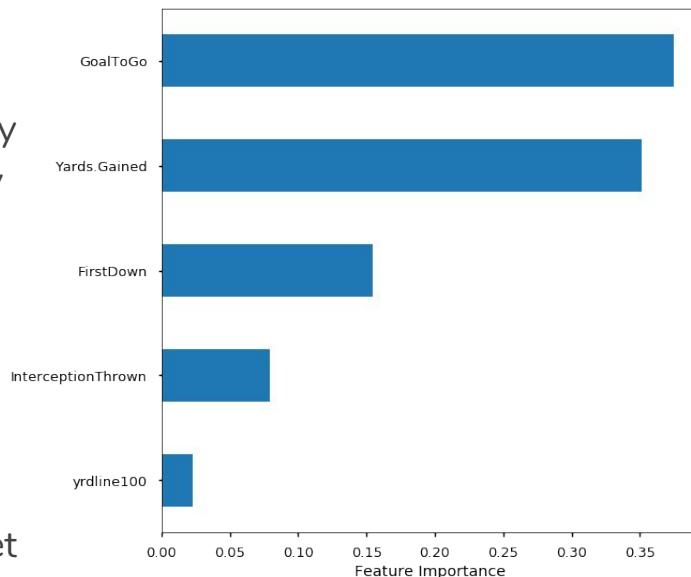
- Train data group = 75%; Test data group = 25%
- Machine learns from training data group
- Uses what it learned from training data to make predictions on test data
- Model used: Decision Trees
- Orange boxes = Non-touchdown
- Blue boxes = Touchdown





Results

- Good at classifying touchdowns; even better at classifying non-touchdowns
 - True negative = 93% accuracy
 - True positive = 89% accuracy
 - Overall testing accuracy = 92.51%
- Accurate enough for reliable & reproducible results, but has enough leeway to make predictions for new data that may be entered into the dataset at a later time





Most Important Features

- Positive Features
 - 1.) Yards Gained
 - 2.) Goal to Go
- Negative Features
 - 3.) First Down
 - 4.) Interception
 - 5.) Yard Line

Touchdown	
Touchdown	1
Yards.Gained	0.011
GoalToGo	0.531
FirstDown	-0.586
InterceptionThrown	-0.205
yardline100	-0.654



How can we apply this information?

- Offensive Coaches
 - Don't get carried away trying to make big plays
 - Close to end zone = higher chance to score
 - Limit turnovers - especially interceptions
- Defensive Coaches
 - Maintain controlled aggression
 - Force interceptions
 - Limit positive yardage by opposing offense



Future Works

- Explore play-by-play data for specific teams
- Explore more advanced modeling techniques & compare results
- If available, explore data from older decades



Thank you! And go Vikings!

- Any questions?

