**NFL 4th Down Analysis**

**Motivation**

* A common idea on “Analytics Twitter” is that NFL teams leave value on the table by refusing to go for it on 4th down. Football is referred to as “A Game of Inches” and we’re reaching a point where every edge a team can get, not just schematically but also strategically, is being explored. I used decision trees and random forests to zoom in on some of the causes and effects of 4th down decisions.

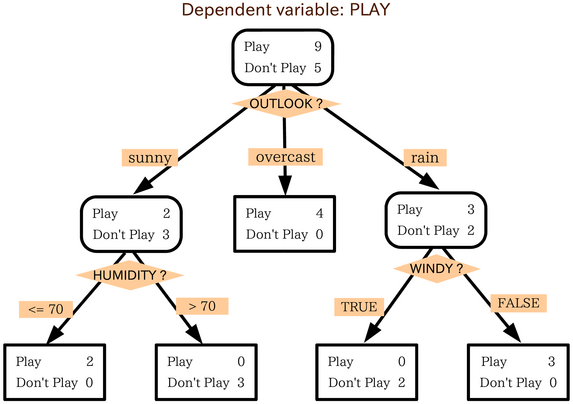
**Description**

* This project uses a [Kaggle database](https://www.kaggle.com/maxhorowitz/nflplaybyplay2009to2016) containing every regular season play from the 2009 – 2016 NFL seasons. This provided over 30,000 4th down plays examine. Multiple decision trees and random forests were built to examine the factors that go into coaches’ 4th down decisions. Situational outcomes were also examined using expected points added (EPA).

**Methods**

* Decision Factor Analysis
* A decision tree algorithm was used to predict whether a given fourth down would result in a punt, field goal, or going for it with 89.9% accuracy.
* The factors the algorithm used were time left in the game, field position, yards to the first down, whether it was a goal-to-go situation, and score differential.
* A random forest algorithm was then implemented and increased the accuracy up to 92.6%
* The weight of the deciding factors from the algorithm were then plotted to see what coaches generally base their decisions on.
* Decision Results Analysis
* After building out the model that can accurately predict coaches’ behavior, it was time to examine the value lost/gained by these decisions.
* Decisions made in various areas of the field (including what I’ve labeled the “Deadzone”) were examined and the average EPA was calculated for different situations.
* The Deadzone is from the 50 yard line to the opponents 35 yard line, where field goals are inefficient, and coaches often punt rather than going-for-it.
* After identifying the situations with the most lost value, another random forest algorithm was used to check out the deciding factors and shed light on the traps coaches fall into.

**Visuals**



* Here is an example of a generic decision tree. The Decision Tree algorithm I used absorbs the chosen features and chooses the most efficient way to label the situation based on past NFL behavior. That could be into a punt, field goal, or go-for-it. The real decision tree I generated contains hundreds of nodes and is too large for a digestible graph.
* A random forest generates a large amount of separate decision trees (in the case of this project 1,000) and uses a voting system to choose a label. This resulted in better accuracy in this project.

A screenshot of a cell phone

Description automatically generated

* Here’s the graphic for the random forest that predicts decisions on 4th quarter 4th downs. You’ll notice field position is the most important, but other factors play significant roles.

**Installation**

* Any version of python past version 3 will work.
* Additional addons required can be found at the top of the nfl.py file
* The csv of the database used is found in the data folder. You must download it to run the program.

**Sources**

The data was used from Kaggle user [Max Horowitz’s](https://www.kaggle.com/maxhorowitz) database. It can be found here or in the data folder.