**MACHINE LEARNING IN PROFESSIONAL FOOTBALL**

Enzo Adriano, Charles Campisi, Ian Valentik, & Nick Weaver

The proposal is to use DVOA data compiled by Football Outsiders and game score data to attempt to predict the winners of the games against the Las Vegas spread and straight-up. We will use historical data from 2015 to 2018 to build the model and update the data with this season’s numbers to increase the prediction value.

One of the difficulties of testing is to find a good benchmark for performance. If our prediction has an accuracy of 50%, is it good, bad or mediocre? It likely outperforms random guessing of win or loss (with ties occurring in less than 1% of games), but is that acceptable? Standard platforms suggest a win rate of 54% against the spread must be achieved to be profitable.

One of the reasons we have chosen spread data is that spread data does not need to incorporate the varying positives and negative rates of the money line, but does need to measure margin of victory and we will use DVOA data to predict the likelihood of a team either beating the spread (the favorite winning by more than the spread amount) or covering (the underdogs losing by less than the spread amount or winning outright).

As an explanation, we will use the DVOA data from previous games to predict future games. We will predict two different results. The first prediction will focus on picking the straight winner (not using the spread, but simply determining which team will win). The second prediction will focus on picking the winner against the spread. This will need to account for the varying benefit of being the underdog (+ points) and the varying negative of being the favorite (- points) when selecting which team is projected to win.

**PROJECT PLAN:**

* Extract Historical and Present Spread/Score Data from <http://www.aussportsbetting.com/data/historical-nfl-results-and-odds-data/>
* Extract DVOA Data from <http://www.footballoutsiders.com>
* Data cleaning will be done in Jupyter or Zepl
* Database storage and management will be done in Postgres/AWS
* Machine Learning
  + Determination of Style
  + Testing the Model
* Tableau will be used for visualizations
* Data Sharing will be done in Colab and Github
  + Github will have Individual Branches per Language and Individual Branches per Individual

**ASSIGNMENT OF WORKLOAD:**

* ETL for Historical and Present Spread/Score Data
  + Charles Campisi
* ETL for DVOA Data
  + Enzo Adriano
* Data cleaning will be done in Jupyter or Zepl
  + Charles & Enzo
* Database storage and management will be done in Postgres/AWS
  + Charles & Enzo
* Machine Learning
  + Nick Weaver
* Tableau will be used for visualizations
  + Ian Valentik
* Data Sharing will be done in Colab and Github
  + Responsibility of all group members to maintain clean sheets, branches, and master