**Drafting a Running Back in the NFL**

As a professional sports team, an NFL organization would like to be able to select players in the annual NFL draft that will be able to immediately make an impact for their team. The goal of this project is to attempt to analyze what factors a potential draft choice running back must possess in college in order to be considered worth being drafted. It is our goal to provide a tool with which NFL teams can determine which players deserve extra attention by their scouting offices. In order to create our model, we are going to rely on data from two different sources.

The first source of data is the NFL itself, which has helpful tables for us to copy and paste into Excel spreadsheets and turn into useful CSV files that contain our data. The tables we are interested in are the ones with NFL running back data from 2010 to 2016. For an example of the 2016 data, follow this link: <http://www.nfl.com/stats/categorystats?archive=true&conference=null&statisticCategory=RUSHING&season=2016&seasonType=REG&experience=&tabSeq=0&qualified=false&Submit=Go>

The second source of data will be from a website called the Football Database. This site helpfully has information on college football players from 2009 to 2015, and we are also able to convert the tables given into CSV files. One issue we may run into with this set is that it only provides information on the top 100 running backs for each given year. Normally this should not be an issue, as it is reasonable to expect that the draft choices will come from the top 100 running backs from the previous year. However, using the link provided above reveals that Rob Kelley on the Washington Redskins team provides an exception to this rule. At this time, it is not known if he is the only exception, but it is something that we will keep an eye on to determine if we need a more robust data source. For an example of the 2015 data, follow this link: <http://www.footballdb.com/college-football/stats/stats.html?yr=2015&conf=&mode=R>

In order to create this tool, we are going to first transform the data into CSV files and check it for missing data. From there, we will combine the college data when available to provide career statistics for the players. In its current form, the data only gives the statistics of the selected year and not for the careers of the players; however, some players show up in multiple years, meaning we can determine some of their career statistics. After we have created the career statistics, we can then use those with machine learning algorithms and our NFL player information from the corresponding years to create a model which will allow us to pass judgment on a set of incoming college running backs.

When we are done, we will have the code from our analysis, a paper explaining our methods and findings, and a model which can be applied to future data sets of college running backs that will determine if they are worth scouting for the upcoming NFL draft.