Predicting rookie season performance from college metrics for NFL quarterbacks
Cole Sutton and Eric Thomas
Level of interest: 10/10

Description:
We explore the predictive power of college quarterback performance statistics on rookie season performance in the NFL. In particular, we predict a rookie NFL quarterback's Quarterback Rating (QBR) based on data from their final year of college play. Given the enormous amount of metrics by which players are measured, it is hard to know what information is important in determining who will end up being an All-Pro player and who will be out of the league in five years. We hope our model will benefit scouts and general managers as they determine which quarterbacks to draft.

Features the data set will include:
The dataset will include but not be limited to the following variables:

	Explanatory variables					**Outcome**
Quarterback	**Average passing yards per game**	**Average rushing yards per game**	**Average touchdowns per game**	**Average interceptions per game**	**Average AP ranking of school during last season**	**Average QBR during rookie season**
Will Ferrell	**305.6**	**34.2**	**3.2**	**0.4**	**17.5**	**67.3**

We hypothesize that the most important metrics for predicting average QBR rating during a quarterback's rookie season are average AP ranking of the school during their last season and passing yards and touchdowns per game.

Some models we will consider using to predict NFL rookie performance from college statistics include the linear perceptron model, multiple regression, decision trees, and the multi-layer perceptron model.

How and from where the data will be gathered:
Sports data are very well documented and go back decades. There are many online websites that we will use to gather the data. Some of the different websites will include sports-reference.com, nfl.com, and espn.com. Different websites have different categories of statistics, so we will want to look at multiple websites to be able to get the best data on a given player for our project.

To gather the data, we will take a random sample of NFL quarterbacks within the last 10 years, and the data on their NFL and college career statistics. Sites such as sports-reference have easily accessible csv files to download of NFL and college data, so we will go through and download data on 100 randomly chosen NFL rookies and then combine the data into a single csv file to be trained.