STT450-550 Project Proposal

Patrick Connell, Skylar Furey, and Zack Licursi

We are looking at data of concussions in the NFL and their many impacts. We are first looking at how a player getting injured with a concussion in a game impacts the outcome of the game, win or lose. We know that a player’s position and average play time impacts the influence a player’s injury has on the result of the game. We will also look into which positions are more likely to get hurt, and how well different teams do in coaching their players how to reduce concussions. We can also see how these change over the years and how much of the variance is based on randomness. We can also test how getting injured affects the players playtime before and after the injury. Further we can see if players who miss more games have more affected play time. There are 517 entries and nine variables.

Player- The player that got injured during the game. This is a supervised categorical variable and will be used in classification analysis.

Team- The team that the player who got injured played on at time of injury. This is a supervised categorical variable and will be used in classification analysis.

Position- The position that the player who got injured played during time of injury. This is a supervised categorical variable and will be used in classification analysis.

Winning\_Team- Whether or not the injured players team won the game. This is a supervised binary variable and will be used in classification analysis.

Season- Which season the game was that the player got injured. This is a supervised categorical variable and will be used in classification analysis.

Games\_Missed- How many games the player missed after the game the player was injured in. This is a supervised numerical variable and will be used in regression analysis.

Playtime\_Before- The number of downs the player played in games on average before suffering the injury. This is a supervised numerical variable and will be used in regression analysis.

Playtime\_After - The number of downs the player played in return game after suffering the injury. This is a supervised numerical variable and will be used in regression analysis.

Playtime\_Lost- The number of downs less the player plays in their return game than they did on average in games before there injury. This is a supervised numerical variable and will be used in regression analysis.

I am concerned with how little data points our set has with only 517. This concerns me because it allows skewness to impact analysis more than in a big data set. I am happy that there is an even mix of positions with only full backs and defensive backs lacking in volume. The years unfortunately aren’t nearly as even with there being much less injuries reported in 2014/2015 than the other years.