**Predicting NFL Plays for Optimum Success**

**Related Literature/Industry Research and Outcomes**

In the NFL, a single play-call can be the difference between a win or a loss. A designed run play may work well in a certain situation, while it may work terribly in a different situation. Therefore, it is vital to make use of analytics to know *when* and *where* these calls need to be made. The NFL is fairly new to analytics, so you often see new innovations and projects geared towards making insightful decisions. For example, Kaggle hosts an annual NFL data scientist competition called the “Big Data Bowl”. (National Football League, n.d.)It is a football analytics competition affording college students and professionals the opportunity to utilize historical data sets of the same player tracking data used by teams and suggest innovations about how football is played and coached. The winners are invited to Scouting combine to present their project to coaches. From this competition you can find many different viewpoints of analyst around the world and some of those projects gave me the inspiration for my own project. Some of my inspiration also came from social media as well. There’s a large community of sports analysts who interact and encourage new ideas. Ron Yurko (Twitter, n.d.)(who’s nflscrapr package has been used in many varieties) and Micheal Lopez (Director of Analytics for the NFL) are prime examples.

**How this project is Different**

There’s been a lot of NFL related projects, such as predicting run play success (2020 Big Data Bowl), pass play success or even concussion related projects. There’s been many variations of “predicting the opponents next play”, but a lot of the models I’ve seen have had poor accuracy scores; hovering around 30%. My job and project is to significantly improve on these accuracies and produce a score around 70-80%.

**Transformation, Data Cleaning and Exploratory Data Analysis**

Before using my dataset, I had to transform and clean it. It had many unnecessary columns, relating to specific play and player personnel. So, I selected the ones I would need for my models to predict the next play type. There were missing values in those valuable columns as well, so the next step was removing rows where this was the case. I then performed Exploratory Data Analysis on the dataset. The Redskins have 13 distinct opponents in the 2020 season, so I will eventually create 13 separate datasets for each team to build models upon. For Delivery 2, I focused on the Redskins’ division rivals: Dallas Cowboys, Philadelphia Eagles and the New York Giants. I compared the number of passing attempts to rushing attempts per each down from 2009-2018 and further analyzed to see what area of the field these passes are being thrown to. This can give a coach a general idea of where to focus their defense on each down when playing against these teams. I will also embark on further analysis, such as analyzing the rushing attempts as well. Will also be looking at each plays’ WPA (Win probability added) and EPA (Expected Points added) in comparison with other teams.

Williams Anosike

# References

National Football League. (n.d.). *NFL Operations*. Retrieved from Big Data Bowl: https://operations.nfl.com/the-game/big-data-bowl/

Twitter. (n.d.). Retrieved from Ron Yurko: https://twitter.com/Stat\_Ron/