**Executive Summary** – A complete but concise description of your work to entice potential readers into reading the remainder of the report. It should answer the following questions:

• Why should the reader care about the problem and results?

• What problem is the report going to try to solve?

• How did you go about solving the problem?

• What are the results?

• What are the implications?

Intro 1: Defensive Coordinators in the National Football League make millions of dollars leading their team’s defense, hoping to allow them to be victorious. Although Defensive Coordinators are usually seasoned veterans that are very good at looking at the current situation and inferring what the other team may call, it is almost an educated guess every time the defense calls a play.

Intro 2: One of the biggest battles in a football game is not between any of the players on the field, but between the Offensive and Defensive Coordinators. The defense is trying to run the most optimal play to stop the offense, while the offense is trying to run the most optimal play to make the desired yardage, and ultimately a touchdown. Although Defensive Coordinators are usually seasoned veterans that are very good at looking at the current situation and inferring what the offensive team may call, it is almost an educated guess every time the defense calls a play.

We have found a way to take the idea of play calling and enhance it from an educated guess an individual makes to a model that can predict the offensive play call with approximately 70% accuracy, which takes the guessing game out of football. Our model takes different play metrics; down, yards to go, side of field, etc. to predict what the Offensive Coordinator is going to call. In order to solve the problem, it was important to determine which factors were most influential in predicting the next play that would be called. We determined that factors were most significant in predicting the play call. After using a random forest model that takes in all of the important factors, the model tells us whether or not the team will run; left or right, or pass; short, middle, or long. The model was able to take data from 2013, and correctly guessed the next play going through the 2014 data with a 75% prediction average.

Since our model has the ability to correctly predict the next play called 75% of the time, this model could be very useful to a Defensive Coordinator who is trying to call the best defensive play to stop the offense from gaining yards, but more importantly keep them from getting a first down, and ultimately a touchdown.

**Motivation and Dataset Overview** – Why did you pick this problem? What is the problem you are actually trying to solve and why did it interest you. A clear problem statement is key because it will focus all of your data wrangling and analytical efforts.

Here is also a good place to provide a little background information about your topic. Be sure to include at least two outside sources (with proper citation) to help provide context for your problem and possible avenues of analysis.

The dataset overview should provide a short description of the dataset you used to conduct the research. At a minimum, you should address:

• Where the data came from – a specific source should be cited

• What type of data it is (cross-sectional, time-series, panel)?

• How was the data collected?

• Provide a list or table of variables used in the analysis and how they are defined (don't forget units).

• What are some basic descriptive statistics (e.g., number of observations, number of variables, what each observation represents, summary statistics if appropriate, density/sparsity of the data)?

The dataset overview should make it clear to the reader where the data came from and allow the reader to get the data in order to re-create your analysis. In total, the motivation and dataset overview section should set the stage for the rest of the report, giving the reader a clear picture of what the problem is, why the problem was important to you, and the data you will use to address the problem.

We think that football analytics is coming along and going to be much more prevalent in the near future, and this will be one of the ways it will be used. We have heard that Gary Kubiak, the head coach of the Denver Broncos’, is trying to replace one of his coordinators on the headset with an analyst that will be running regressions and using statistics to determine the next play that should be run (Wagner-McGough). One of the very interesting analyses online is from the Green Bay vs. Seattle NFC Championship Game, in which Brian Burke breaks down multiple decisions throughout the game and statistically shows how each decision changed the outcome of the game (Burke).

The purpose of the project is to be able to input the data points of the current play; the yards to go, the down, etc. and predict what the offense is going to run. If the Defensive Coordinator was able to use analytics in order to determine what the offense is going to call and ultimately what he should call in order to stop this. This would take all of the guessing out of play calling and get rid of any bias the Defensive Coordinator has; if the defensive team is most likely to blitz on 3rd down, the offense could then look at that and call a play that would combat that.

The data came from an online source by Bryan Povlinski in which he collects play by play data from every game, all 256, where he collects over 58 variables. The data is given in Excel spreadsheet and is compatible with Stata. We have data from the 2013 season and 2014 season, but so far the analysis has focused on the earlier year’s data. However, exploring our data we have come to the conclusion that about 22 of the variables will most likely be explanatory in our model, which are listed in the Appendix A. The many variables can be cut down further with more exploration. Another 4 are related to the dependent variable, as they have to do with what the team’s play call is and where the ball is going.

The type of data we have acquired is cross-sectional data. While there is a time variable present in our data, we do not see it significant to test how variables differ over time. We are looking at many individual subjects, in this case offensive plays from the NFL regular season, therefore not making time-series or panel data applicable.

Our data is very complete, especially for its size. Of the almost 2 million data points only 7,693 are not filled in, which is less than 1% of the data. However it appears the missing data is not absent, but represent an action not happening. For example, if there isn’t a fumble on the play there is no data point in the fumble column. This is relevant for other variables like interceptions, receiver, or pass location, which would not have values if the play wasn’t a pass. To deal with this problem we have created dummy variables with a zero when an action didn’t happen and a 1 when it did.

**Conclusions, Implications, and Recommendations for Further Work** – Draw a conclusion about the work you did and how you answered the original problem statement. Describe any implications your work has for the field you researched. Lastly, provide some recommendations for further work; if you had more time, money, resources, what else would you do to better address your problem.

Future Work- DCs, split up model (85% accurate on 3rd down)

The goal of the project was to predict the next play the Offensive Coordinator will call given the current play parameters. Our model predicts the outcome correctly 68% of the time in terms of run versus pass, but only 45% correct prediction when trying to predict the specific play (run right, pass short, etc.). Although our model did not have a prediction accuracy percentage that seems astounding, our model was still able to predict the next play. The model is covering such a wide range of circumstances, such as quarter and leading or losing, so 68% accuracy for run versus pass with all of these specific circumstances not accounted for is good; especially considering individuals that split up the data by quarter and team have only gotten an average of 75% correct (ScienceDaily).

As we talked about earlier, Gary Kubiak of the Denver Broncos is very interested in play analytics. He would like to replace one of his coordinators on the headset with an analyst in order to help Kubiak make real-time decisions. Our model solidifies the idea that analytics in sports is increasing in popularity, and could become very prevalent in the near future.

We reached out to the Air Force Football Defensive Coordinator, Steve Russ, because we were hoping to receive some statistics from him that he may know as a football coach, and once we explained our project, he was very interested in what we were studying. He wants us to reach out to him once the football season has ended so we can talk about our model with the hopes of him being able to apply it to college football. If we had more time to work on the project, we would split up the model into certain specific situations; quarters, teams, and leading or losing. Splitting the model up may allow us to be more accurate because football coaches may be more willing to run earlier in the game, or when they are leading, and pass later in the game, or when they are losing. Specific teams also play a big role in the willingness to run or pass because a team that has a dominant quarterback, like the New England Patriots, are going to be much more willing to throw the ball than a team with a strong run game. There would be many ways to make the model more accurate because of the expansive ideas and play options teams have, but our model does a very good job at predicting the next play a team will call.

**Appendix**

1. **Variable Present in Data**



\*Highlighted variables were originally chosen by us to include in our analysis, the bottom row represents the outcomes we wish to predict

Works Cited

American Statistical Association. "Statistical model predicts with high accuracy play-calling tendency of NFL teams." ScienceDaily. ScienceDaily, 12 August 2015. <www.sciencedaily.com/releases/2015/08/150812103645.htm>.