Correlation Between College Football Player Statistics, NFL Combine Results, and NFL Draft Pick Ranking*

Jonathan Adams[1] and Julian Swart[2]

***Abstract*— The purpose of this analysis is to determine if a correlation exists between a college football player's collegiate success on the field, National Football League (NFL) combine test results, and his resulting NFL draft pick. The analysis is broken down by player position due to the varying nature of available variables at each position. The collegiate statistics will include relevant game metrics such as yards gained, passes caught, completions, tackles, and various others. The NFL combine test results include metrics such as bench press repetitions, forty yard dash time, vertical jump, and more. The NFL draft pick is ranked based on the order in which a player was selected by NFL teams. The resulting outcome of this analysis will be to provide college football athletes a clearer picture of what metrics matter the most to increase their draft equity. This will allow the athletes to focus their time into appropriate training.**

I. INTRODUCTION

The NFL is the pinnacle of success for an American football player. Hundreds of college football players in America work hard towards the dream of making it to the NFL and playing for their favorite team. However, an extremely small percentage of collegiate level athletes make it to the professional level. The NFL draft is yearly event in which each of the thirty two current NFL teams come together and decide which college athletes to add to their roster for the upcoming football season. Several factors play into a team's decision on which players to draft. A few of these factors include: current team roster needs (at which positions are they in need), player success on the field in college, NFL combine test results, and player attitude. This analysis will focus on tangible results obtained from collegiate in-game statistics and combine test results. The NFL combine is also a yearly event in which top performing college players are invited to come participate in workouts and drills while recruiters from NFL teams watch and evaluate their performance. A few of these tests which we will be analyzing are the 225 pound bench press repetition test, the forty yard dash, the vertical jump, and the three cone drill. These tests are all designed to determine a player's physical abilities. While watching the NFL combine, the commentators often talk about how a certain player has increased his draft ranking based on his performance on these tests. The following analysis will attempt to determine if a player can affect his draft stock (positively or negatively) based on NFL combine performance metrics.

*This work was not supported by any organization

[1]J. Adams is a Graduate Student in the Business Analytics Program, University of Tennessee, 201 Stokely Management Center, Knoxville, TN 37966

[2]J. Swart is a Graduate Student in the Business Analytics Program, University of Tennessee, 201 Stokely Management Center, Knoxville, TN 37966

II. PROCEDURE FOR DATA COLLECTION AND CLEANING

Data for this analysis will be collected from several different sources. The NFL combine test results will need to be obtained from This website has historical combine results dating back to 1987. The collegiate statistics for each player will be collected from This website has statistics for all relevant positions dating back to 2004. The NFL draft pick ranking will be obtained from The draft results website contain archives dating back to 1936. Due to the limited historical college statistics data, this analysis will only include data from 2004 - 2017.

The data will need to be scraped from these websites. After the data has been obtained, it will definitely need to be cleaned up. There are missing test results on the combine website and not every player who is drafted attends the combine. The data will also need to be merged into one table for analysis. The technologies of choice for these tasks will be Python and R. Python will be used for the data collection and cleaning. R will be used for merging the tables.

III. DATA ANALYSIS

After gathering all the cleaned data into a clean table, our analysis will be done using R. The analysis will be done at various levels. The initial approach will be to attempt to find correlations between the combine results, college in-game statistics, and the resulting draft rank from the entire data set. The next step will be to break the analysis down into groups based on individual position. There are vast differences between in-game metrics and physical attributes between each position on a football team. Therefore this level of analysis will most likely be the most useful.

The initial attempt will involve trying to determine a player's individual draft rank. However, factors outside of the data we have, such as the current team's positional needs, may affect the specific slot that a player gets drafted. Therefore, the analysis will most likely force draft ranks to be put into buckets such as draft round or draft rank buckets such as positions 1-5, 6-10, etc.

IV. DESIRED OUTCOME

The desire outcome of this project is to generate analysis that determines which aspects of a football player's resume is most useful in increasing draft pick rank. Along with the analysis of which factors play the greatest role in draft

success, a hopeful outcome will be to create a predictive model that given in-game statistics and combine results as input would return a predicted draft ranking.

If sufficient time allows, another hopeful outcome would be to predict success in the NFL based on the combination of data we have already collected. However, this would be a last step if everything goes smoothly.

A. Milestones

- Obtain the data from all the various sources
- Look through data and determine necessary cleaning steps
- Merge tables together to form one table for analysis
- Do initial exploratory data analysis (aggregating, summary statistics, etc.)
- Begin model building and validation for prediction of draft rank
- Test models using subset of with held data

B. Hopeful Milestones

- Obtain NFL game statistics for players in the data set
- Repeat the cleaning and exploratory steps
- Begin model building and validation for prediction of NFL success