ISyE/CSE 6740 - Spring 2021
Final Report

Student Name: Ved Mohan
Project Title: Predicting Cornerback draft order in the NFL

Introduction and Literature Review

Since 1936, like many professional sports leagues, the National Football League (NFL) has implemented the annual reverse order draft. This allows the worst performing teams a chance to pursue the best incoming talent, theoretically re-balancing the distribution of talent overtime. Each year, franchises are hard pressed to evaluate each draft class' talent, in order to estimate production at the next level.

On the other hand, players are similarly motivated to improve their demand or "draft stock". In a study of the yearly compensation of NFL quarterbacks and running backs, Simmons & Berri (2009) found that players drafted the earlier rounds can merit higher pay for the entirety of that individual's career.

Wide receivers (WRs) are an essential component of an offense in an increasingly pass-heavy league. Often the fastest individuals on the field, they have been the focus of many academic papers, but there has been less of a spotlight on their defensive counterpart, Cornerbacks (CBs). CBs are expected to make the same high velocity cuts and match WRs, step for step.

In a competitive environment such as the NFL, understanding the factors that contribute to a high draft pick are of utmost interest to NFL hopefuls. The intent of the project is to **combine collegiate affiliation information and physical measurements to predict draft viability and rank.**

There are two main components to this paper-

1. What makes a successful draft candidate?
2. What separates the top 10 CBs from the rest of their draft class?

Methodology

Data

Data is collected on a total of 400 CBs that attended the NFL combine during the 2010-2021 seasons. Of these 400, 273 athletes were successfully drafted. Each year, athletes dedicate hours to train for the NFL combine. They participate in a variety of standardized drills and tests to prove their readiness to compete at the next level. The description of the combine statistics used in this report are shown in Table 1. The sample statistics are contained in Table 2.

College Pedigree

Following the lead of numerous researchers (Treme and Allen (2011), Mullholand and Jensen (2014), Fenn and Berri (2018)) a dummy variable (Pedigree) was created. This variable takes a value of one if an athlete

Variable Name	Brief Description
Drafted	1 if the athlete was drafted, 0 if not
Overall Draft Pick	Overall draft pick number, -1 if not drafted
Age	Athlete's age in years
Height	Athlete's height in inches
Wt	Athlete's weight in pounds
Vertical	Vertical jump distance in inches
40YD	40 Yard dash time in seconds
BenchReps	Number of bench press repetitions
Broad Jump	Horizontal broad jump distance in feet
3Cone	3 Cone agility drill time in seconds
Shuttle	Shuttle agility drill time in seconds
Conference	College football affiliation of an athlete's school
Pedigree	1 if Conference has a proven record and reputation, 0 otherwise[1]

Table 1: Description of features used

Variable Name	Count	Mean	Standard Deviation	Minimum	Maximum
Drafted	400.00	0.68	0.47	0	1
Overall Draft Pick	400.00	72.95	74.98	-1	254
Age	339.00	21.80	1.04	14.00	24.00
Height	400.00	71.34	1.60	67.00	76.00
Wt	400.00	193.43	8.80	169.00	220.00
40YD	381.00	4.50	0.09	4.28	4.75
Vertical	325.00	35.77	2.66	29.00	44.50
Bench Reps	321.00	14.41	4.12	2.00	26.00
Broad Jump	325.00	122.44	5.54	109.00	147.00
3Cone	247.00	6.92	0.19	6.28	7.55
Shuttle	254.00	4.17	0.14	3.82	4.58

Table 2: Sample Statistics

played in large "powerhouse" conferences, and zero otherwise. The list of the conferences includes the ACC, Big 12, Pac-12, Big 10, SEC.

Imputation of missing values

Some CBs choose not to participate in certain combine drills such as the broad jump, the vertical leap, etc. As shown by Enders (2010) the missing data can be estimated using the Multiple Imputation method based on a Multivariate Normal distribution. The iterative imputer package from Sklearn was used, which is built on multiple imputation by chained equations as its underlying methodology (Buck 1960). Due to the Bayesian assumption of this imputation, analyses were aggregated over 100 different samples to account for randomness.

Evaluation

Combine metrics and their effect on being drafted

Initial analysis of combine metrics was performed using both logistic regression and Lasso Regression[2]. These models were selected to identify the significant features, the results of which are shown below in Table 3. The metrics with the most impact on being drafted are the 40 yard dash, weight, the 3 cone drill.

Variable	**Logistic Regression Model**	**Lasso Regression Model**
Age	-0.14	0.00
Height	0.03	0.00
Weight	**0.56**	**0.07**
40YD	**-0.70**	**-0.11**
Vertical	0.02	0.00
Bench Reps	-0.01	0.00
Broad Jump	-0.08	0.00
3Cone	**-0.24**	**0.006**
Shuttle	0.14	0.00
Pedigree	-0.14	0.00

Table 3: Logistic and Lasso Regression coefficients. Models were cross validated and tuned, with significant variables in bold.

This observation directly parallels research done by Fenn & Berri in 2018, who found that the single most influential factor in a WR being drafted was the 40 yard dash. Understandably, CB success is determined by their ability to keep up with their direct counterparts. This is further enforced by the selection of the 3 cone drill. This drill evaluates how fast athletes can change direction while accelerating, which is essential in blocking and intercepting passes. The inclusion of weight reinforces that athletes must not only be fast, but also carry the mass required to make physically demanding blocks and tackles. This reduces the chance of CBs being "mis-matched" and being taken advantage of by large WRs, such as Seattle Seahawk's D.K. Metcalf (See Fig 1.).

Figure 1: Seattle Seahawk's wide receiver D.K Metcalf (229 lbs.) matched up against San Francisco 49ers' cornerback Emmanuel Moseley (184 lbs.)

Decision tree analysis was performed to validate the initial regression findings. Decision trees were made using both gini impurity and information gain to test the stability of the thresholds, and were found to be consistent. Decision trees were selected over random forests to maintain interpretability. The feature *Age* was the selected threshold on multiple levels (Fig 2.). While initially surprising, this has been explored in

[2]Data was standarized for this analysis using SkLearn's Standard Scaler package

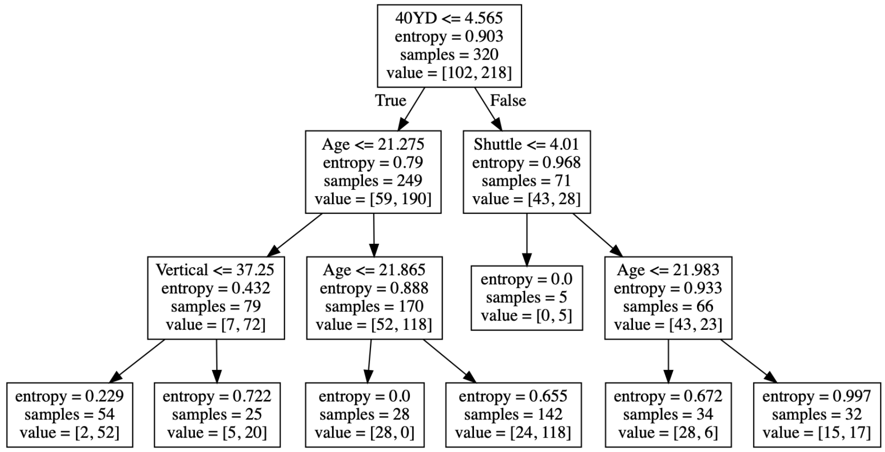

Figure 2: Decision tree with depth set to three. *Age* was used as a threshold three times. There was no significant appreciation in accuracy at a depth of four (accuracy = 0.76 vs 0.753) so a depth of three was maintained for simplicity of interpretation.

previous research. J. Mulholland and S.T. Jensen (2014) found that press and news coverage of athletes positively related to the likelihood of their getting drafted. Players will only feel confident in declaring for the NFL draft if they are talented, creating a self fulfilling cycle of younger players declaring for the draft and getting selected.

To explore the impact of purely physical performance on the draft, player age was removed as a factor. With a reduced accuracy of 0.70, the new decision trees reflect *two* different ”prototypes” of CBs: lithe and speedy, and mass-heavy and slower. The initial threshold is made on the 40 yard dash time, affirming both the initial regression analysis and the intuition that CBs are drafted to match WRs. Subsequent splits are made on *Height, Shuttle, Bench Reps*, and the *40 Yard dash.* These features can be broadly classified into ”Speed/Agility” and ”Weight” categories respectively.

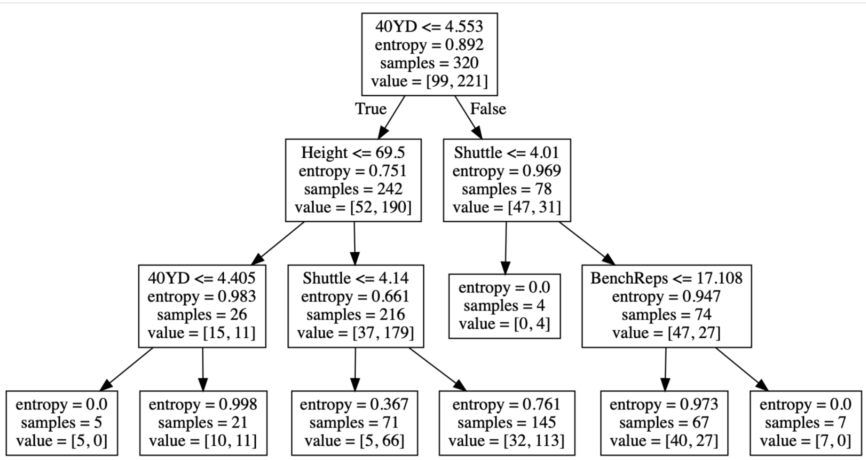

Figure 3: Decision tree with depth set to three and *Age* removed as a feature. There was no significant appreciation in accuracy at a depth of four (accuracy = 0.68 vs 0.69) so a depth of three was maintained for simplicity of interpretation.

Separation between players in the same draft class

To understand the factors that create differentiation between players within the same year, they are to be evaluated within scope of one another. The overall draft pick order of athletes were processed into a relative ordering (Eg. In 2016, Athlete 1 from FSU is picked 5th overall, Athlete 2 from Ohio State is the second cornerback picked, 10th overall. Therefore, relative within CBs, Athlete 1 is ranked 1, and athlete 2 is ranked 2).New derived metrics were created in line with the methodolgy followed by Berri and Simmons (2011), and Berri and Fenn (2018). See Table 4. for a description of the features used.

Variable Name	**Brief Description**
Weight	Athletes weight in pounds
BMI	Weight divided by height (used to capture muscle mass)
Jump	Broad jump and vertical jump added, in inches
Agility	3Cone, Shuttle, and 40 Yard Dash summed, in seconds

Table 4: Description of derived features used

K-means clustering with K=2 is used to divide the class into two "prototypes", defined by the cluster centers for the derived statistics. The most frequent prototype within the top 10 relative rank is compared to the most frequent protype within the remainder of the same years picks. This captures the factors typical of relative ranking label within draft-classes. As can be seen in Table 5., there is a clear separation. The top 10 CBs each year are on average leaner and jump higher than the remainder of the draft class. This implies that currently, NFL teams prioritize reachability and agility of CBs rather than their mass.

Year	BMI	Weight	Jump	Agility
2010	-1.66	-15.00	1.34	0.00
2012	-1.44	-14.88	-2.89	-0.07
2013	-1.30	-16.10	-0.93	-0.21
2014	-0.45	-14.86	-1.26	-0.09
2017	-0.65	-18.29	7.96	-0.12
2018	-2.38	-16.73	1.30	-0.37

Table 5: Difference between the first 10 prototype CBs selected and the rest, using K means clustering. On average, CBs that are selected early are leaner, jump higher, and are more agile, relative to the cornerback class for each year.

Conclusion and Future Work

In summary, the primary novelty of this project is the study of a new position that has no supporting literature, the Cornerback. It was found that cornerbacks are broadly able to be placed into two classes, smaller, faster athletes, and larger, slower athletes. Furthermore, the smaller, faster CBs are preferred are seen as more desirable, and constitute the majority of the top 10 picks, relative to the position itself. This implies that the league and its talent scouts emphasize the reach and mobility of CBs rather than their physicality.

An observation that was a departure from previous literature- collegiate pedigree did not seem to influence the draft stock of Cornerbacks. Since CBs in the "blue-blood" schools often match up against future NFL WRs, it may prove to be a disadvantage if they are regularly out played. Atlanta Falcon's own A.J. Terrell (playing for Clemson, from the ACC conference) garnered negative press due to his mis-match against the Cincinnati Bengal's Ja'Marr Chase (playing for LSU, from the SEC conference), which has been cited as his fall from the number one cornerback in his year (LB Sports 2020). This analysis is extremely situational and game specific, so match up statistics will have to be analyzed on a game by game, player by player basis.

Data Sources

- https://www.sports-reference.com/cfb/
  - College affiliation
- https://stathead.com/
  - CBs height, weight, shuttle cone drills, vertical and horizontal leaps, etc.

Github Repository Link

https://github.com/ved-mohan/ISYE6740

References

- Berri, D. J. and R. Simmons. 2009. "Catching A Draft: On the Process of Selecting Quarterbacks in the National Football League Amateur Draft." Journal of Productivity Analysis, September 18, 2009.
- Buck, S. (1960). A Method of Estimation of Missing Values in Multivariate Data Suitable for use with an Electronic Computer. Journal of the Royal Statistical Society. Series B (Methodological), 22(2), 302-306. Retrieved May 4, 2021, from http://www.jstor.org/stable/2984099
- Enders, C. K. (2010). Applied missing data analysis. New York: Guilford Press.
- Fenn, A. J., Berri, D. (2018). Drafting a Successful Wide Receiver in the NFL–Hail Mary? International Journal of Sport Finance, 13(1), 18–33.
- https://larrybrownsports.com/college-football/aj-terrell-burned-jamarr-chase-clemson-lsu-national-championship-game/533117
- Mulholland, J., Jensen, S. T. (2014). Predicting the Draft and Career Success of Tight Ends in the National Football League. Journal of Quantitative Analysis in Sports, 10 (4), 381-396. http://dx.doi.org/10.1515/jqas-2013-0134
- Pitts, J. D., Evans, B. (2019). Drafting for Success: How Good Are NFL Teams at Identifying Future Productivity at Offensive-Skill Positions in the Draft? The American Economist, 64(1), 102–122. https://doi.org/10.1177/0569434518812678
- Simmons R, Berri DJ (2009) Gains from specialization and free agency: the story from the gridiron. Rev Ind Organ 34(1):81–98. doi:10.1007/s11151-009-9200-9