Predicting NFL Performance from Combine Results

Jake Wojcik

Problem & Context:

The NFL Scouting Combine is an annual event where college football players perform physical and mental tests in front of National Football League coaches, general managers, and scouts. The purpose of the combine is to provide insight into the medical history, athletic abilities, psychological state, as well as, demonstrations of skill in positional drills of potential prospects. It's long been wondered do the results of these physical tests have any effect on future NFL Performance? The goal of this project is to predict individual players NFL performance based on Combine results.

Target Clients:

The primary clients this project targets are NFL coaches, scouts, and general managers. The clients spend millions of dollars on the acquisition of talent every year when drafting players. The results of this project will hopefully aid them in picking potential players in the NFL Draft.

Data:

The dataset used for this project was created by Sports Viz Sunday (https://data.world/sportsvizsunday) and posted on data.world. The dataset comes in a single excel file with two sheets: one for NFL Combine results (10,228 rows, 16 columns) and one for NFL Draft results (8,435 rows, 34 columns). I'll be using data from the years 1987 to 2015 in my project. The dataset is available for download via the following link: https://data.world/sportsvizsunday/nfl-combine-data/workspace/file?filename=NFL+Combine+Data.xlsx

Data Definitions:

Columns	Description
Player_id	Unique player identifier with draft year followed by name.
Draft_Year	Year player was drafted.
Rnd	Round number of the NFL Draft player was selected.
Pick	Pick number of the NFL Draft player was selected.
Tm	Team that drafted player.
Player	Name of player
HOF	Whether player was selected to the Hall of Fame or not
Pos	Position of player
Position Standard	Position of player
First4AV	First 4 years approximate value of player
Age	Age of player
To	Last year player competed in the NFL
AP1	Number of All-Pro selections
PB	Number of Pro Bowl selections
St	Number of years player started for their team
CarAV	Career approximate value
DrAV	approximate value accumulated for the team that drafted the player
G	games played
Cmp	Pass completions
Pass_Att	Pass attempts
Pass_Yds	Pass Yards
Pass_TD	Passing Touchdowns
Pass_Int	Passing Interceptions
Rush_Att	Rushing attempts
Rush_Yds	Rushing yards
Rush_TDs	Rushing Touchdowns
Rec	Receptions
Rec_Yds	Receiving yards
Rec_Tds	Receiving touchdowns
Tkl	Tackles
Def_Int	Defensive interceptions
Sk	Sacks
College/Univ	College or university player attended
Height (in)	Height of player at NFL Combine
Weight (lbs)	Weight of player at NFL Combine
Hand Size (in)	Hand Size of player at NFL Combine
Arm Length (in)	Arm length of player at NFL Combine
Wonderlic	IQ test score
40 Yard	40 yard dash time
Bench Press	Number of bench press reps completed
Vert Leap (in)	Vertical leap in inches
Broad Jump (in)	Broad jump in inches
Shuttle	The player starts in a three-point stance and then runs five yards to his right, touches the ground, reverses and runs back 10 yards, touches the ground, before heading back five yards to the finish line
3Cone	Designed to measure speed, agility, change of direction, body control among other traits.
60Yd Shuttle	Same as shuttle but 60 yards instead of 20.

Target Variable

Data Wrangling:

The main issue with the raw dataset was missing values and the need to merge the two separate sheets. Because I could only use players who were both drafted and attended the NFL Combine, after merging the two datasets, the resulting data frame was 5643 rows and 51 columns. I removed the 'HOF' column due to all of the values being 'no' which was clearly incorrect. I also converted the 'sk' column into a float because it was previously an object data type. I dropped unnecessary columns including the 'To' column. Instead I replaced it with a new feature by subtracting 'To' from 'Draft_Year' to get the number of seasons played.

As for dealing with missing values, since dropping rows with any missing values would reduce the dataset to almost nothing, I decided to impute the missing values. Many rows had Na values for Combine results due to players bypassing certain drills for various reasons. The easiest and most efficient way to fill these missing values is by using K Nearest Neighbors (KNN) method. KNN is an algorithm that is useful for matching a point with its closest k neighbors in a multi-dimensional space. It can be used for data that are continuous, discrete, ordinal and categorical which makes it particularly useful for dealing with all kinds of missing data. The assumption behind using KNN for missing values is that a point value can be approximated by the values of the points that are closest to it, based on other variables.

Exploratory Data Analysis:

I looked at the correlations between every feature and the target variable (Approximate Value). The goal of this project is to predict NFL performance based on Combine results. Therefore, choosing a response variable for NFL success must take into account several factors, as not all success in the NFL looks the same. I believe the best response variable is career approximate value: a method of putting a single numerical value on any player's season, at any position. This numerical value takes into account a player's position so it doesn't favor skill positions like quarterback or receiver. Ultimately, it measures the overall value the player had for his team, which I believe is the greatest measure of success for a team sport like football. You can find more information on how approximate value is calculated here:

https://www.sports-reference.com/blog/approximate-value-methodology/

Key Findings:

The first noticeable takeaway while exploring the data was we see a large majority of players have unsuccessful NFL careers, as most players hover around the 0-10 career AV range (Figure 1). Over 40% of all players have a career approximate value of less than 10. This distribution follows closely that of the Pareto distribution: a heavy-tailed distribution that is sometimes used to model distribution of incomes. The basis of the distribution is that a high proportion of a population has low income while only a few people have very high incomes. This is the same for NFL success, only a small portion of players become wildly successful in the NFL while most players experience very little success.

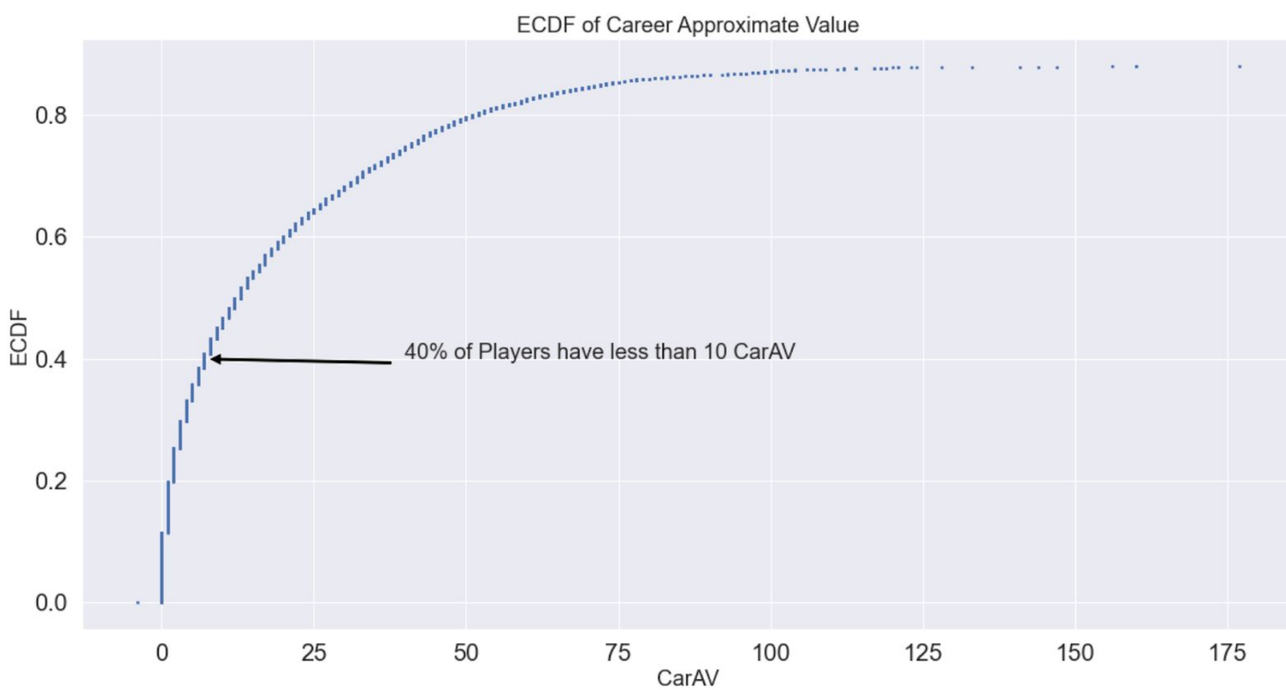

Figure 1: Empirical distribution funtion of career approximate value.

The second key takeaway was the relationship between draft pick and approximate value. Figure 2 shows higher picks generally have more success in the NFL for all positions excluding punters, but there is very little data points for punters drafted. Generally, punters and kickers are picked up as free agents. Although, there are plenty of examples of players picked lower in the draft that had phenomenal success in the NFL. Tom Brady is one great example. He was picked in the 6th round 199th overall. He is clearly visible in figure 2.

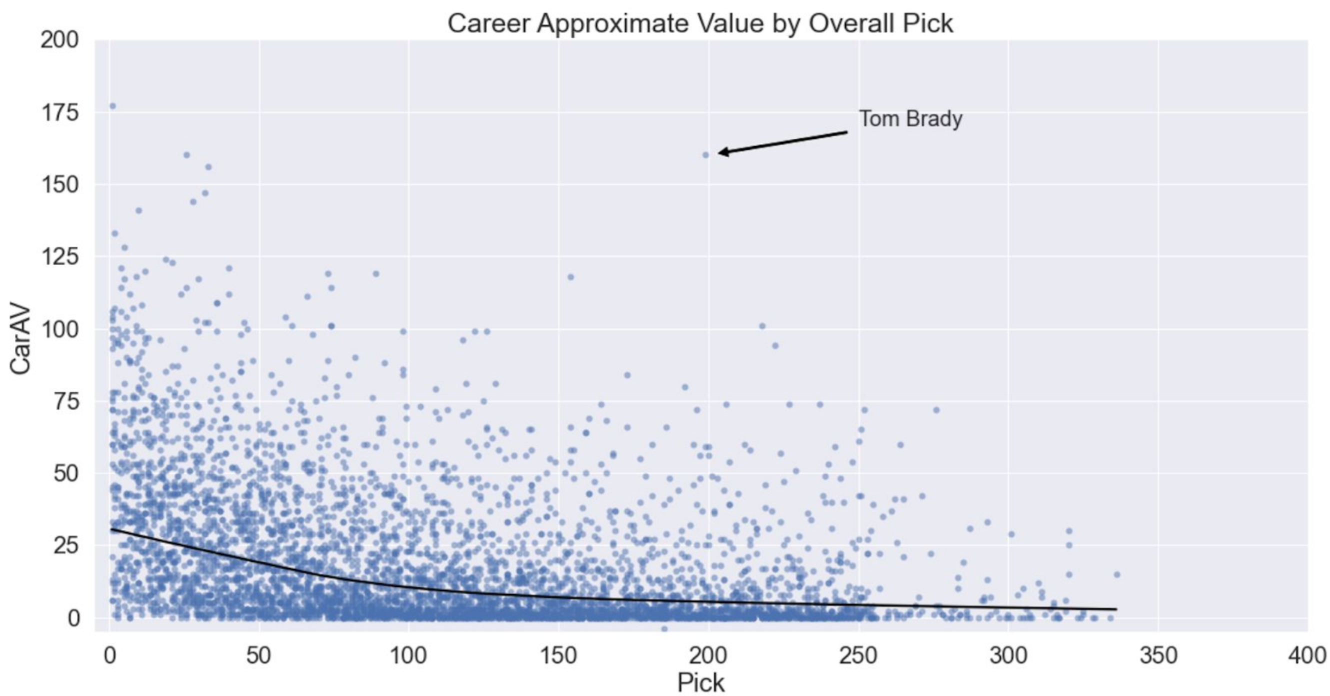

Figure 2: Career approximate value by overall pick.

Hypothesis Testing:

Since 40 yard dash time is widely regarded as the most important and most watched metric at the Combine, I wanted to explore it further by asking a simple question:

- Does a Player's 40 Yard Dash Time Impact Where a Player is Drafted?

- There is a linear shaped relationship between skill position players' (WR, RB, DB) forty times and draft position (Figure 3). Implying the faster the 40 time, the lower the draft pick.
- We can test this hypothesis by permuting pick feature values but leave the 40 time values fixed. This simulates the hypothesis that they are totally independent of each other. Then, for each permutation we will compute the Pearson correlation coefficient and assess how many of your permutation replicates have a Pearson correlation coefficient greater than the observed one.

- Result: Out of 100,000 permutations not one of the Pearson correlation coefficients was larger than the observed one. It is extremely unlikely that 40 times doesn't have an effect on draft position.

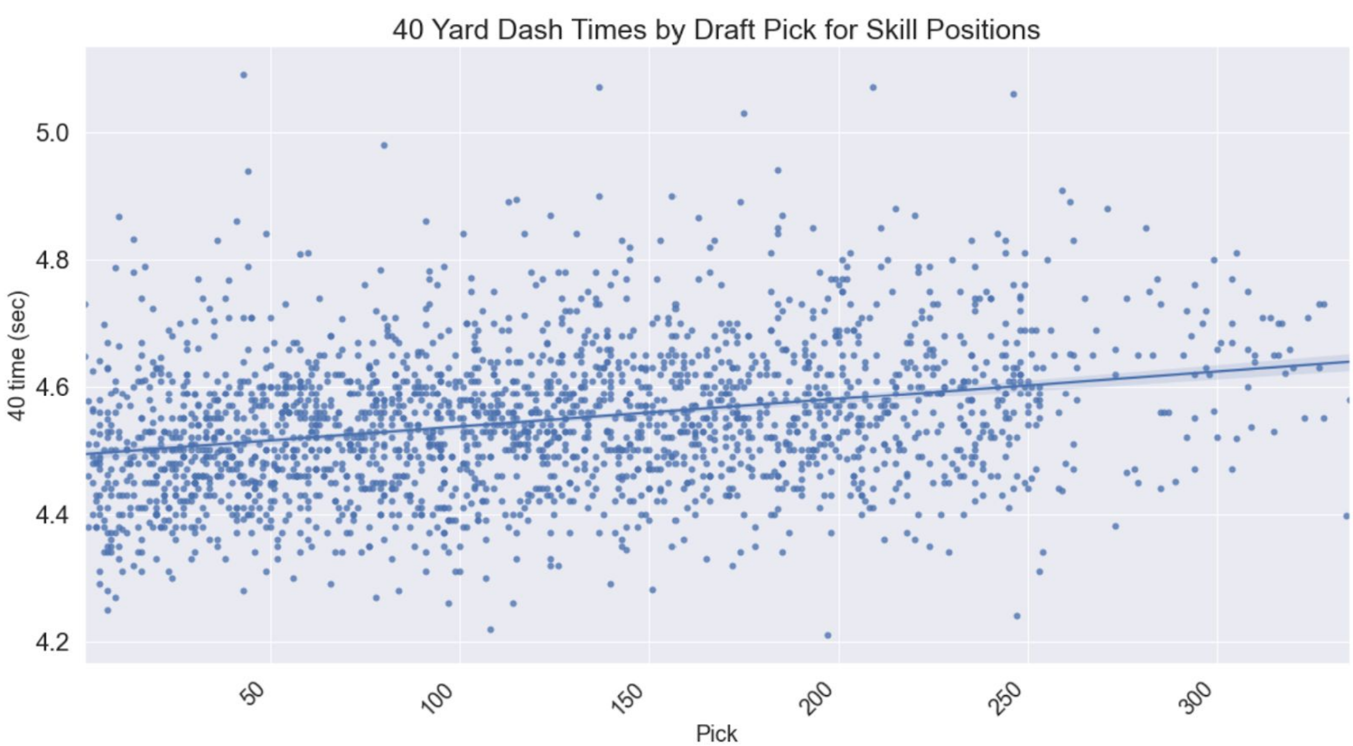

Figure 3: 40 yard dash by draft position for skill players.

Modeling:

I chose to work with the Python scikit.learn library for training and testing models. I chose root mean squared error (RMSE) and adjusted R^2 as evaluation metrics. For hyperparameter tuning, I used 5 fold cross validation using scikit-learn's grid search method. Since the target variable is continuous, I used supervised learning and regression models on the data.

Before modeling, I split the train and test data manually to prevent data leakage. The training data will be all drafts from 1987 to 2010, and the test data will be the last three drafts, 2011 through 2015. The training and test data account for ~80% & ~20% respectively. Ultimately, we want our model to predict future results. By splitting the data in this way we can improve the models' ability to generalize to new data.

Before trying to predict NFL success from the Combine, I wanted to see what features contributed to the target variable the most. I set up a simple random forest regressor leaving in all the career stats features and reviewed the most important features. The "Pick" feature was the most important feature by far, which I thought was interesting given all the career stats were used to train the model. Now that the relationship between draft position and NFL performance is established, I attempted to examine how the Combine results and the draft order are tied together. If we can show that having good Combine results have positive effects improving one's chance of being drafted early or being drafted at all, then we could safely assume that the effects are being carried over to one's NFL performance, since we have already made it clear of the relationship between draft order and NFL performance.

I broke up the modeling into three sections. To determine the best model, I tested 4 models on each section; linear regression, random forest, SVM, and gradient boosting. The modeling sections include:

1. Predict NFL Success from Combine Results

- The Support Vector Machine was the best performing model with a RMSE of 2.
- R^2 values were very low indicating no predictive power at all.

2. Predict Draft Position from Combine Results
    - The gradient boost model had better success predicting draft position than NFL success but R^2 value was still low (.10) and RMSE was off by a considerable amount (67.47).
3. Separate by Player Position
    - Since the Combine is more important for some positions, I decided to separate the dataset by player position to see if the models improved any.
    - Unsurprisingly, quarterbacks had little to no predictive power.
    - I was able to improve the model's score in wide receivers and running backs by using a ridge regression model and tuning hyperparameters.

By separating the players by position I was able to improve models slightly, especially for skill positions. Table 1 shows the modeling metrics for each position.

Position	Adjusted R2	Root Mean Squared Error	Features Dropped
Quarterbacks	0.059411	73.686792	season stats
Wide Receiver	0.090329	66.619473	season stats
Defensive Backs	0.075714	65.233829	season stats
Running backs	0.097121	58.783971	season stats

Table 1: Comparison of model metrics for each position.

We can see there is almost no predictive power in determining draft position for quarterbacks. It is a little surprising to see doing well on the Combine rarely has any effect in increasing one's chance of being drafted higher. Ultimately, what is being measured in the Combine is one's pure athleticism which is always a great attribute to have as an athlete. This might be due to the fact that there are position specific drills for every position, and most of the attributes of NFL quarterbacks are best demonstrated by actually throwing the ball.

A player's pure athleticism is far more important at skill positions than any other position. We would expect to see better model performance among skill positions, and that's exactly what was observed. By far the most important drill in determining draft position for skill players was the 40 yard dash (Figure 4).

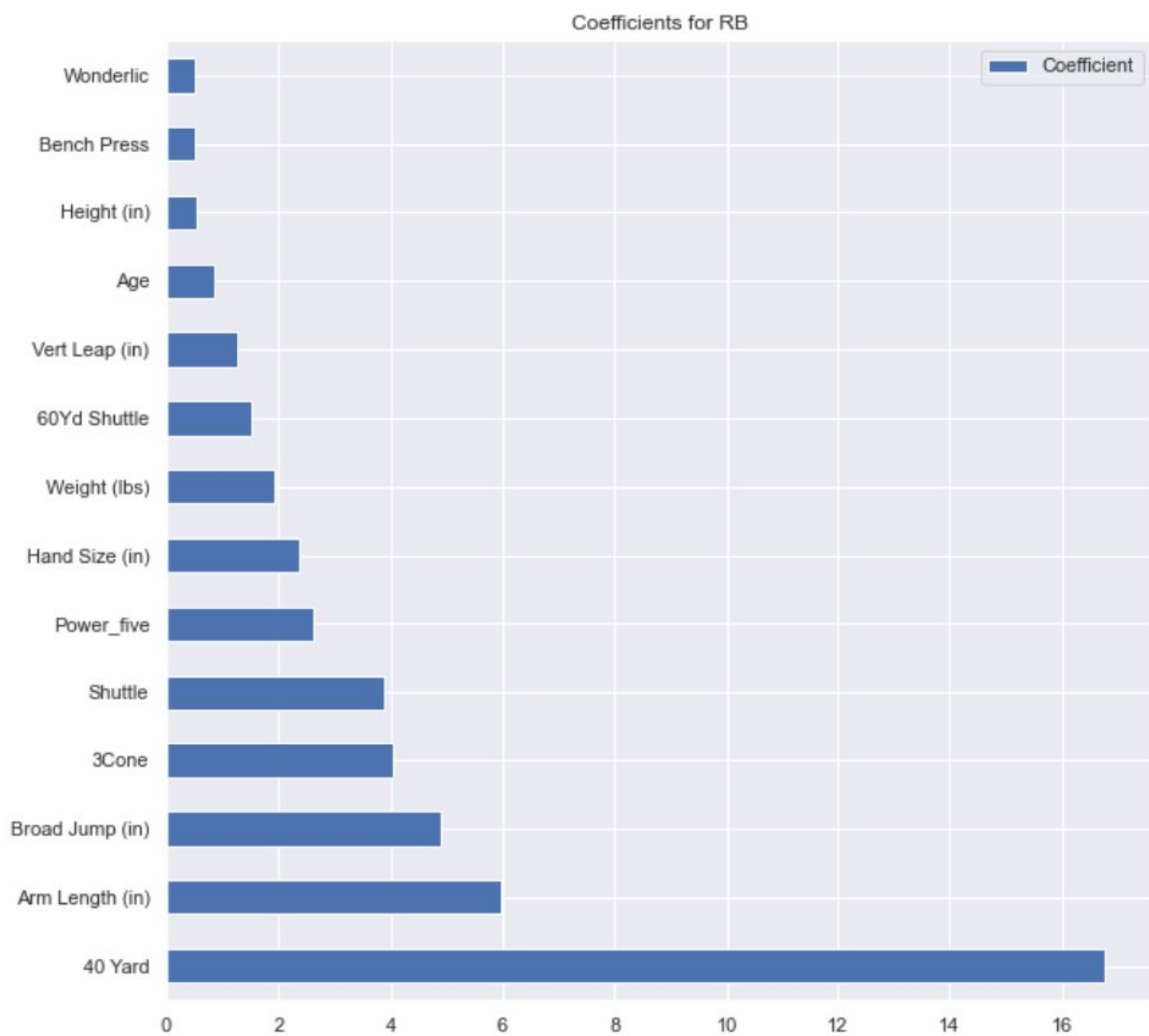

Figure 4: Feature coefficients for ridge regression model among running backs.

Besides the 40 yard dash, the 3 cone drill and shuttle contribute to the models prediction significantly. Both of these drills require agility and acceleration while moving laterally, essential skills for running backs in the NFL. It appears coming from a power five conference in college is somewhat important for running backs, but not important for the other positions modeled. A large part of what makes a running back successful is breaking tackles and sprinting away from opposing defenders. Defenders in power five conferences are generally bigger, faster, and stronger than smaller conferences so the ability for running backs to break tackles and out pace defenders in power five conferences is more impressive than excelling in smaller ones.

Predictions:

Overall the best performing model still has little predictive power in determining draft position and, in turn, NFL success. Predictions were generally off by a fair amount (Figure 5,6). Being off by an average of ~59 picks is very bad considering each round is 32 picks. I would consider this model a success if it were able to predict the round drafted reasonably well, but here it's off by almost two whole rounds.

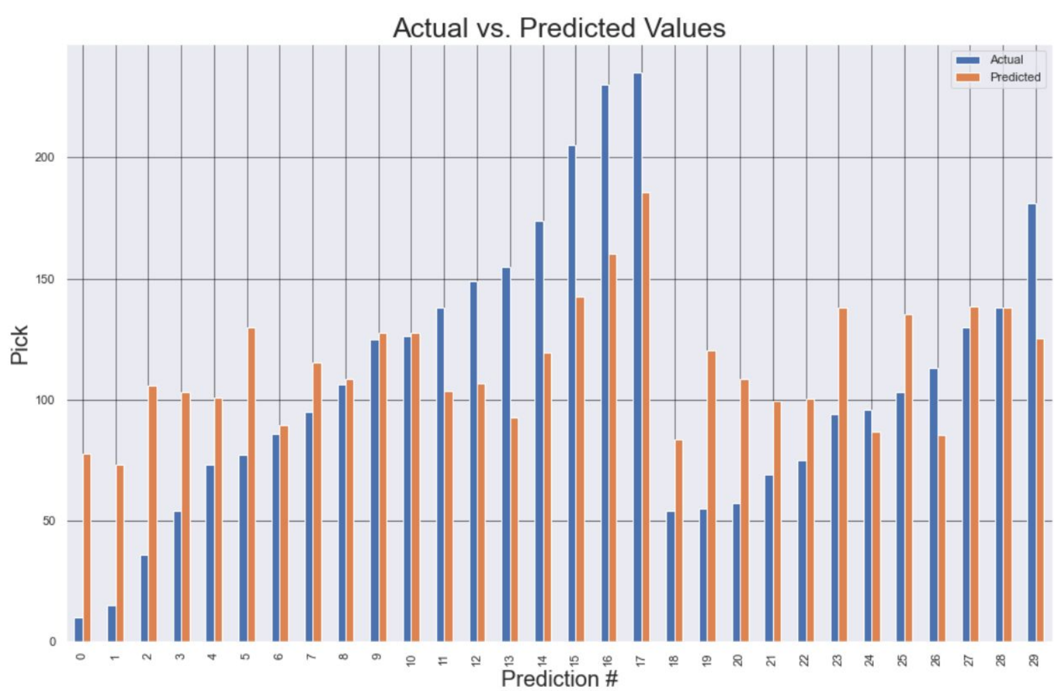

Figure 5: Actual vs. predicted values for running backs using the ridge regression model.

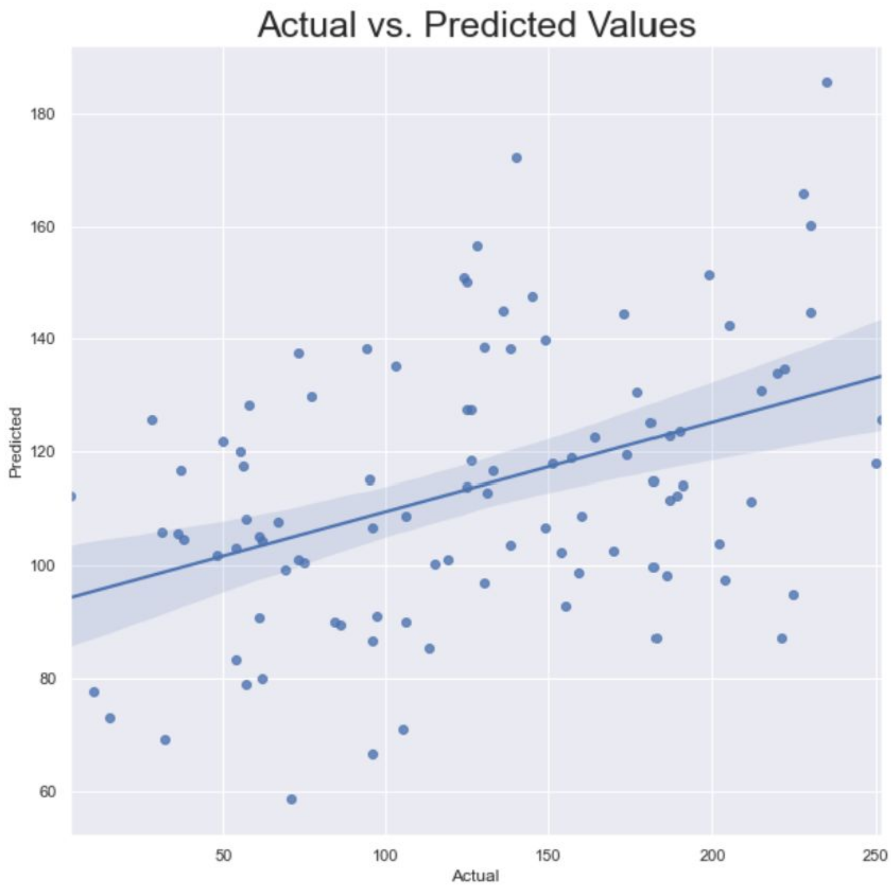

Figure 6: Actual vs. predicted values for running backs using the ridge regression model. As you can see, there is very little correlation between the two variables.

Conclusions & Key Takeaways:

It's surprising that most of the Combine results lack any prediction power for draft position or NFL performance. Raw athleticism may not be as critical as casual NFL fans claim it to be. A player's true talent should reveal itself on the football field rather than at the Combine. So does the NFL Combine matter? The scouting Combine isn't only for measuring athleticism but coaches, scouts, and general managers all have the opportunity to interview players to get an understanding of their character. Work ethic, motivation, and attitude *combined* with athleticism are what it takes to really succeed in the NFL. Furthermore, players who are likely to succeed in the NFL are going to get noticed by scouts in college. Traditional scouting remains the best way to determine who to pick on draft day. Scouts consider top prospects long before the Combine which explains why many players who do not perform well in the Combine still get drafted early and go on to have successful NFL careers.

Of course there are exceptions, but there are also exceptions in reverse. Take Vernon Gholston for example, Gholston performed exceptionally well in the Combine. He performed above average in every drill except the 60 yard shuttle and in most cases performed at an elite level (Figure 7). Gholston is a perfect example of a "workout warrior" at the NFL Combine. These are players that have superior measurables and physical attributes who outperform expectations. Many times a good workout at the Combine will improve a players draft stock as scouts are impressed by their superior athleticism. This is exactly what happened with Vernon Gholston. Drafted as a pass rusher that could get to the quarterback, he ended his career with a whopping 16 tackles and zero sacks. Players like Gholston have become warnings to scouts to not fall in love with the athletic measurables because they don't always translate on the football field. And this is the key takeaway for scouts and coaches alike: take the Combine with a grain of salt. Use Combine measurements as supplemental material when selecting a player but don’t put too much stock in them.

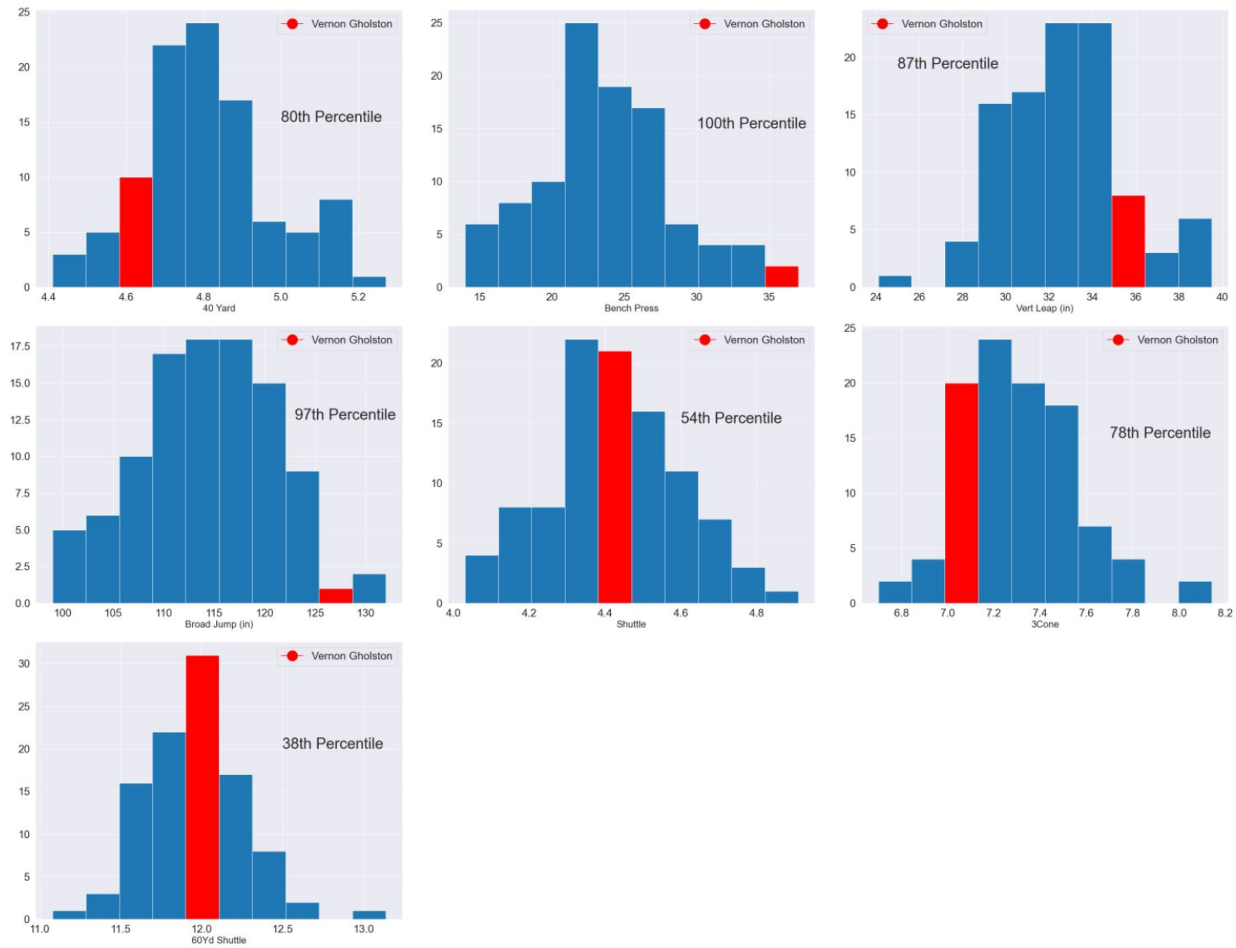

Figure 7: Distribution of Combine results among defensive ends with Vernon Gholston highlighted in red.

Further Reading:

There have been a few other studies that have tried to predict NFL performance based on Combine measurements. The study conducted by Paul Park concluded very little predictive power for all positions excluding running backs ((Park, 2016). This conclusion is similar to my study. The study by McGee and Burkett derived a regression equation with the draft order as the response variable and 11 different Combine measurements. It concluded that the coefficient of determination, namely R2, was significantly high, for positions including WR,RB, DB, and QB indicating players in these positions were more likely to be drafted if they had excellent Combine measurements (McGee, K, and Burkett L., 2013).

Future Research:

Further research utilizing college statistics could improve prediction models for NFL performance for all positions. College statistics would add a whole new dimension to this dataset. It would be interesting to see the difference in NFL performance from players who excel at smaller colleges compared to players who excel at powerhouse programs. It would also be helpful to obtain more data from the most recent draft years and see if that improves the models any. Lastly, I think it would be interesting to create a “draft grade” feature, assigning a grade to every player going into the draft based on scouting and media sentiment at the time.

References

Park, P. (2016, August). Does the NFL Combine Really Matter. https://www.stat.berkeley.edu/~aldous/Research/Ugrad/Paul_Park.pdf

McGee, K, and Burkett L. The National Football League Combine: A Reliable Predictor of Draft Status? Journal of Strength and Conditioning Research. 17(1): 6-11, 2013.