Understanding Agility

ABSTRACT

The NFL Combine (the “Combine”) is an annual event used to assess college football players entering the NFL draft through a series of athleticism tests. Using data from the Combine, this analysis aimed to use linear regression to interpret how strength, speed, explosiveness, and body type associate with agility.

DESIGN

Training athletes is a complex process that involves honing many attributes in order to make a player maximally effective on the field of play. Furthermore, you will find very little consensus among strength coaches on the optimal way to do this. Agility is especially complex because it involves combining many athletic skills. It is easy to associate squatting more with a higher vertical jump but agility is far more difficult to unpack. This presents an opportunity for data science to provide strength coaches with insights into attributes associated with agility.

TOOLS

Beautiful Soup (Web Scraping); Pandas & Numpy (Data cleaning, manipulation and analytics); Seaborn and Matplotlib (Visuals); Scikit Learn (Regression modeling); Stats Models (Summary statistics)

DATA & ALGORITHMS

Data was acquired through web scraping pro-football-reference.com. I scraped the last twenty years of Combine results using Beautiful Soup. The original data from the Combine contained *position, height, weight, fourty yard dash time, vertical jump, broad jump, three cone drill time, and shuttle time*. Important engineered variables included *vertical jump ÷ broad jump* (“vert/broad”) and *height ÷ weight* (“lankiness”). Vert/broad was an effort to capture how vertically explosive a player was relative to horizontal explosiveness. Lankiness is a measure of body type that was calculated and named in a way that would make it more obvious and intuitive to interpret.

The y variable for measuring agility was shuttle time. While the three cone drill is also a measure of agility, I felt shuttle was a better event mainly for its simplicity relative to three cone drill.

It is important to note that all numerical features used in the modeling process and referenced hereinafter were standardized using the mean and standard deviation within that position group.

Because of the interpretive nature of this project, I built several simple models and compared the results. Models were trained and tested on an 80/20 split. The train and test outputs were used for summary statistics and visuals but every model was run through cross validation in order to examine variance in testing results. LASSO was used to determine if any features should be removed. I also used LASSO to determine which categorical dummy to drop.

Key Findings

In a group containing the positions (see glossary) DB, LB, and DE, a **one standard deviation increase (relative to position group) in vert/broad was associated with 0.04 seconds faster in shuttle time**. This occurred in a model with categorical dummies for position and vert/broad as features. The important takeaway here is the *direction* of the vert/broad coefficient.

This indicates a higher vertical power relative to horizontal power is associated with faster shuttle time. Mean absolute error (“MAE”) and R squared on the test data were 0.13 seconds and 0.23, respectively.

Running a model identical to the previous except replacing vert/broad with forty time, shows **one standard deviation faster (relative to position group) in forty time was associated with 0.04 seconds (rounded) faster in shuttle time**. Forty time and vert/broad intend to measure two different attributes that both play into the shuttle drill and their coefficient absolute values show to be almost identical in this case. MAE and R squared on the test data were 0.13 seconds and 0.25, respectively.

When running the same models on a group of only the position OL, **one standard deviation higher (relative to position group) in vert/broad was still associated with 0.04 seconds faster in shuttle time and one standard deviation faster (relative to position group) in forty time was associated with 0.10 seconds (rounded) faster in shuttle time.** In both of the OL models, MAE on the test data was 0.14 seconds (rounded). The predictive accuracy gets far worse in the OL models, but the vert/broad coefficient remains negative. Due to the nature of the position, there are far more outliers among OL.

R squared was mainly used to assess over/underfitting and for comparing very similar models. MAE was much more insightful for interpretation. From my knowledge of the shuttle drill, I believe 0.15 seconds to be a tolerable upper limit for MAE. The models I have discussed have MAE that is certainly high enough to be material in assessing a player, but it is still small enough to take results as meaningful. In addition to R squared and MAE I used the stats models package mainly to explore standard errors of the coefficients and p values. Overall, the predictive accuracy of the model is less important than the consistently observed negative value in the vert/broad coefficient and its similarity to forty time as a feature in the DB, LB, DE group.

Bench press (adjusted for body weight) and height (relative to weight) had no meaningful association with faster shuttle times. This does not necessarily mean that these variables have no association with running speed or change of direction. For example, shorter players could have quicker change of direction but be slower in straight line speed and the effects cancel out, leaving shuttle time unaffected by height. The lack of insights from these variables raise questions for further exploration.

COMMUNICATION

See presentation slides and Jupyter notebook containing code and visuals.

Glossary of Positions Referenced	
DB	Defensive Back
LB	Linebacker (Defense)
DE	Defensive End
OL	Offensive Line