In the 1990s almost no one was as popular and well-known of an athlete as Hulk Hogan. Hulk was everywhere; he was on television wrestling, he was on talk shows, and he was starring in action/comedy movies. Hulk Hogan had become a household name. Imagine if you could predict Hulk Hogan’s popularity back in the 1970s when he debuted. That would be amazing! As a wrestling fan it likely would not lead to financial gain, however, what if the wrestling promoters could have predicted Hulk’s popularity for monetary gain?

Who is the Hulk Hogan of the NFL? Some might argue that Tom Brady, quarterback for the New England Patriots and Tampa Bay Buccaneers may be considered a household name. He has appeared in numerous commercials and movies reflecting his pop culture status as a household name. He is also one of the most decorated players in history earning ten (10) All-Pro designations. What if executives and coaches in the NFL could predict who the next Tom Brady will be? This is what I aimed to do with this project.

In the National Football League (NFL) there are certain players who stand out for their on-field performance. Each season accredited media members, NFL players, and NFL executives vote on who are the best players at each position for both offense and defense. These patchwork hypothetical teams are called the “All-Pro” teams. Since All-Pro teams are voted on by the groups listed, they are considered to be more prestigious than the fan-selected “Pro Bowl” teams. Also, All-Pro players are considered the best at their position regardless of their team’s overall performance. This focus on individual performance instead of overall team performance allows for extremely talented individuals to shine and be recognized even if their team’s win-loss record does not reflect their skill.

It also stands to reason that the more All-Pro players a team has playing for them, the more successful the team will likely be and therefore generate more revenue. From a fan-perspective this is beneficial as your favorite team will play better and provide bragging rights. From an executive perspective, like attracts like, meaning the more talented players you have (All-Pro) then other talented players will want to join them and multiply the team’s success. The goal of this project is to be able to use NFL Combine performance to predict who may become a future All-Pro.

The NFL Combine is a multiday general workout for prospective NFL players to participate in and hopefully garner enough attention to be drafted. There are multiple events and activities that the prospects compete in. For the purposes of this project each prospect’s performance was measured in multiple exercises. There are agility & speed drills such as the 3 Cone drill, Shuttle Run, & 40 yard sprint. The Bench Press is the only strength drill included. There are also jumping drills included in the Vertical Jump and Broad Jump. Prospect names, positions, height, weight, and collegiate alma mater are also included in the data.

NFL combine performance and All-Pro data were gathered from pro-football-reference.com via web scraping. The combine performance data made up the bulk of the dataset while the number of times a player was selected to an All-Pro team was added as the target variable. Over 900 players since 2000 have been selected to All-Pro teams, with many being selected multiple times from multiple teams.

A screenshot of a cell phone

Description automatically generated with low confidence

After combining the number of All-Pro designations (labeled “AP Count”) with the combine performance metrics, the dataset resembled the table below:

A screenshot of a graph

Description automatically generated with low confidence

Some players did not have performance listed in some of the drills. For example the first row above, for John Abraham there were NaN values for Vertical, Bench, Broad Jump, 3Cone, & Shuttle. In cases like these I chose to replace the NaN values with a zero instead of removing the datapoint entirely. This way the zero did not alter the impact of other players who did have performance in those drills. Some players did not have a Height or Weight listed either. In those cases I chose to utilize the mean value of the respective metric. These players obviously have a height and weight, it was not included for some reason and therefore using the mean would allow other variables to influence the model more. After solving for NaN values, the player names were dropped. Next dummy variables were generated for position played and school attended. Lastly, the numerical values were standardized via StandardScaler. This was done so that incremental changes in each metric were weighted the same.

A picture containing text, screenshot, font, number

Description automatically generated

Above is a correlation matrix of the numerical variables in the dataset. We can see that there are some strong correlations between certain variables that are likely not a surprise. Height and weight are highly correlated, which makes sense as the taller a player is, the more likely they are to be heavier. Performance-wise, there are strong correlations between weight and bench press, indicating that the more a player weighs the stronger they are likely to be. It is also interesting to note that the agility drills such as 3Cone and Shuttle correlate highly with the jumping drills of Broad Jump and Vertical.

In order to account for multicollinearity, I decided to drop the features of 3Cone, Shuttle, and Broad Jump. These are all highly correlated with other features and may muddy the results. While height and weight are also highly correlated, I felt it was important enough to keep both metrics.

I also chose to keep the AP Count target as an ordinal integer between 0 and 10. This is because many players, like Tom Brady for example, had numerous All-Pro designations. I wanted players with multiple All-Pro selections to have their Combine performance weighted higher than players with only one or two. Then the data was split into 75/25 training/testing datasets and entered into a Logistic Regression model. After the model was trained and tested on previously unseen testing data, it had an accuracy score of 88.8%. From an odds-ratio perspective, the most impactful features were playing specific positions:

* Offensive Line (1.35)
* Defensive Line (1.29)
* Edge (1.28)

The most impactful schools were:

* Hawaii (1.29)
* Washington State (1.27)
* Louisiana Tech (1.27)
* Auburn (1.27)

While 88% is a highly desirable accuracy score, it likely does not tell the entire story of a player’s ability. There are many variables unable to be accounted for; injuries are unpredictable, off-the-field issues with law enforcement, performance-boosting substances that are banned, and other issues that cannot be mathematically predicted. There are also other human factors outside of the players themselves. The coaching and executive staff of the team they get drafted to have a huge influence. If a coach’s game plan does not highlight that player’s skillset to the best of their ability, then they may not shine in the spotlight as they may in another game plan/scheme. Players are also fairly powerless when it comes to which team may draft them. As is the case with any sports league, some teams have more capital or cap-space to spend putting their rosters together than others. Some teams are more frugal and depending on previous performance may be in a “rebuilding” mode where they are accepting losing more games now in order to leverage draft picks and future cap-space for building a better roster in upcoming seasons. Below is a chart showing the number of All-Pro designations since 2000.

A picture containing line, screenshot, plot, parallel

Description automatically generated

The New England Patriots (NWE) have had 103 All-Pro players in comparison to the Detroit Lions (DET) who have had 32. It is reasonable to assume that a player drafted to the Patriots or to the Baltimore Ravens (BAL) will have a better opportunity to become an All-Pro compared to the Detroit Lions or St. Louis Rams (STL).

My recommendation for this prediction via logistic regression is to have NFL executives and coaches utilize this as one of the many tools in their toolbox when considering drafting players. Similar to how wrestling promoters and Hollywood producers saw dollar signs in Hulk Hogan, simply putting stock in every wrestler who has the same build, stature, facial hair, and tan may not yield the same results. There will always be the human element of athletes that cannot be predicted but can be mitigated.

**Q & A**

1. What NFL seasons were included?
   1. 2000 - 2022
2. What is the All-Pro designation?
   1. A player is designated as an All-Pro if they are selected for their individual position via voting from accredited media members, other NFL players, and NFL executives.
3. What is the Pro Bowl team?
   1. Similar to All-Pro, Pro Bowl teams are two full teams of offensive & defensive players selected as the most popular in their respective position.
4. What is the difference between All-Pro and Pro Bowl?
   1. All-Pro selections are voted on by “insiders” in the NFL (players, executives, & media members) whereas Pro Bowl selections are almost exclusively decided by fan voting. While players do need to achieve a certain level of satisfactory play to be nominated to Pro Bowl voting, the player does not need to necessarily be the best player at that position. Also, Pro Bowl players then have the opportunity to play in a game titled the Pro Bowl. This is where the players voted on play a light version of football for fun and spectacle, rather than winning or losing.
5. Do All-Pro players make more money?
   1. While I was not able to find specific information on mean salaries for All-Pro players, I did find a website listing the top highest paid players for each position. As an example I looked at the Quarterback position and five of the top seven highest paid players have had at least one All-Pro selection. For the Running Back position three of the top three highest paid players have had at least one All-Pro selection.
6. How many All-Pro designations have there been since 2000?
   1. 554 distinct players
7. Which position has the most selections?
   1. Wide Receiver has 75 distinct players designated as All-Pro since 2000. Total, Wide Receivers have also had the most number of designations at 154.
8. What college/NCAA programs have yielded the most All-Pro players?
   1. Alabama has yielded 26 distinct All-Pro players since 2000. Second place is Miami (FL) with 20. Third is a tie between Florida State and Georgia at 19.
9. Who has the most All-Pro designations?
   1. Tom Brady (QB), Matthew Slater (P), and Bobby Wagner (LB) all are tied with 10.
10. Why was NFL team not included as a feature in the logistic regression?
    1. Many All-Pro players were selected as All-Pros for multiple teams, creating duplicate data and possibly creating an unbalanced dataset. Also, many All-Pro players were named as All-Pro playing for teams that did not draft them after the NFL Combine, possibly misleading the model.

**Works Cited**

“2022 NFL All-Pros.” *Pro-Football-Reference.com*, www.pro-football-reference.com/years/2022/allpro.htm. Accessed 2 July 2023.

“2022 NFL Combine Results.” *Pro-Football-Reference.com*, www.pro-football-reference.com/draft/2022-combine.htm.

ESPN Staff. “The NFL’s Highest-Paid Players at Every Position: Here’s Who Makes the Most Money, from QB to Punter.” *ESPN.com*, 1 Sept. 2022, www.espn.com/nfl/story/\_/id/34096853/highest-paid-nfl-players-tracking-most-money-guaranteed-per-year-every-position.

“Hulk Hogan: The Man Who Took Wrestling Mainstream.” *The Sporting Blog*, thesporting.blog/blog/hulk-hogan-the-man-who-took-wrestling-mainstream.