Zach Borromeo

Last Edit: March 5, 2022

[Working Title Goes Here]

\*Note\* This is a draft of content for the final report. Will be converted to Latex later.

Abstract:

The National Football League (NFL) has been hosting an annual player draft since the 1930s [ref], where teams in the league select the best young talent from colleges across the United States to play for their football organizations. In recent times, the process of scouting and drafting players efficiently has become a very important part of building championship caliber teams, with a large emphasis put on player analysis. Every year, draft analysis and big boards and mock drafts come out almost weekly leading up to the draft in mid to late April. With that in mind, the purpose of this machine learning project is to produce two different mock drafts: one that is based on recent draft selection weights and the other based on team needs. These two mock drafts will be produced using both historical and current data. These results are meant to both understand the value placed on different draft selections and to predict with moderate to high accuracy the outcome of the first three rounds of the 2022 NFL Draft. With any success, this process can be used on future drafts and allow for better predictions of player selections.

Keywords:

National Football League, NFL, NFL draft, machine learning, mock draft, player analysis, draft analysis, player value

Introduction:

The National Football League (NFL) has been hosting an annual player draft since the 1930s [ref], where teams in the league select the best young talent from colleges across the United States to play for their football organizations. In recent times, the process of scouting and drafting players efficiently has become a very important part of building championship caliber teams, with a large emphasis put on player analysis. Every year, draft analysis and big boards and mock drafts come out almost weekly leading up to the draft in mid to late April.

Definitions:

* National Football League (NFL) – the National Football League, or NFL for short, is one of the four major North American professional sports leagues. The NFL is home to thirty-two teams, divided equally into the American Football Conference (AFC) and National Football Conference (NFC). The NFL is known as the highest level of American Football in the world.
* NFL Draft – the NFL Draft is an annual event that takes places in late April. During this draft, teams in the NFL are ordered based on overall team record and placement in the NFL playoffs in order to determine who selects players first. The team with the worst record selects first while the defending Super Bowl Champion selects last. Teams are allowed to move up and down in the draft order every year if they are willing to trade away selections to other teams for players or their own draft picks.
* Mock Draft – the overall goal of this project. A mock draft is a simulation or prediction of a draft is set to occur. Given numerous factors such as draft order, team needs, and college football player analysis, mock drafts are primarily made based on expert opinion and insider knowledge.
* Team Needs – a list of a handful of player positions that a given team is looking to fill with a draft pick. In most cases, a team’s needs stem from a loss at a position due to free agency, trades, injuries, or a lack of skill at that position. Team needs are sometimes based on expert opinion whereas other times they are based on factual evidence of player losses due to an aforementioned reason.
* NFL Combine – an event hosted by the NFL that allows for college players to showcase their skills in a variety of ways to NFL team executives, coaches, and scouts. Events include the forty-yard dash, vertical jump, broad jump, and various position specific drills.
* Big Board – the name given to a list of players, ranked in order by overall skill, by an expert or group of experts. The big board is meant to show who experts believe are the best players, regardless of position or positional impact on the field of play.

Player Positions:

Below is a list of the positions that will appear frequently in this study along with short descriptions of their roles on the field.

* Quarterback – arguably the most important player on the field, the quarterback (QB) is in charge of running the offense with peak efficiency in order to score as many points as possible. Quarterbacks are the players who hand the ball off to a running back or throw to their wide receivers and tight ends. Players in this position are stereotyped as tall and big with strong arms, but players with smaller stature have seen success. Common names associated with the quarterback position are Tom Brady, Peyton Manning, Patrick Mahomes, Brett Favre, Aaron Rodgers, Russell Wilson, and Josh Allen.
* Running Back – the running back (RB) is the player that is often tasked with taking the ball from the quarterback and running along some predefined path. His goal is to avoid getting tackled and gain progress on the field in the forward direction. Running backs are often viewed as jack-of-all-trades players, as they are often asked to both run and catch the ball. Players in this role are stereotyped as being on the smaller end of the scale in the NFL, with speed and strength being prioritized by coaches. Variations of this position include the half back (HB) and full back (FB). Common names associated with the running back position include Jim Brown, Barry Sanders, Jerome Bettis, Adrian Peterson, Marshawn Lynch, Derrick Henry, and Christian McCaffrey
* Wide Receiver – the wide receiver (WR) is the player most often associated with running along a predefined path and catching a pass from his quarterback. Wide receivers typically are tall and athletic with very good vision and ball catching skills. Speed is not always a top priority, but several of the best wide receivers are incredibly fast. Common names associated with the wide receiver position include Jerry Rice, Randy Moss, Terrell Owens, Calvin Johnson, Stefon Diggs, Deebo Samuel, and Ja’Marr Chase.
* Tight End – the tight end (TE) is another player associated with catching passes along a predefined path. However, where they differ from the wide receiver is in both their physical build and skillset. Often times, tight ends are asked to provide some extra blocking support on the offensive line during running plays, which means they must be able to block linebackers while also being able to run along a path and catch a pass from the quarterback. Common names associated with the tight end position include Tony Gonzalez, Rob Gronkowski, Travis Kelce, and George Kittle.
* Offensive Linemen – the offensive line is split up into five different positions: left tackle (LT), left guard (LG), center (C), right guard (RG), and right tackle (RT). Often times in mock drafts and player rankings, guards and tackles are not broken down by the side of the offensive line that they play on as they are frequently rotated. Typically, guards and tackles are listed as OG and OT respectively, where centers are always C. Occasionally, players will be listed as OL, meaning offensive lineman. These players have spent time in multiple positions, whether it be center and right guard or right and left guard or right guard and right tackle. Often times, offensive linemen are the largest players on the field with excellent footwork and blocking skill as they have the job of protecting the quarterback and making space for the running back. Common names associated with the offensive line include Jonathan Ogden, Orlando Pace, Jason Kelce, Joe Thuney, Nate Solder, and Michael Oher.
* Defensive Linemen – the linemen opposite the offensive line, the defensive line is also made up of a few different positions. Depending on the defensive scheme, players on the defensive line include the defensive tackle (DT), defensive end (DE), and nose tackle (NT). These players vary in skillset, but their main goal is to get penetration up front and push into the backfield where the quarterback and running back primarily play from. Defensive linemen are large players like offensive linemen, with their skillset being comprised of strength, good footwork, penetration, and tackling. Common names associated with the defensive line include Aaron Donald, Chris Jones, Derrick Thomas, and Bruce Smith.
* Linebacker – the linebacker (LB) is a position that has multiple variations. These players are multitalented in what they can do. Often times they are viewed as players who can stop the run, tackle well, and provide some pass coverage towards the middle of the field. Variations to this position include outside linebacker (OLB), inside linebacker (ILB), and middle linebacker (MLB). Common names associated with the linebacker position include Ray Lewis, Lawrence Taylor, Derrick Brooks, Junior Seau, and Brian Urlacher.
* Cornerback/Defensive Back – the cornerback (CB), also known as the defensive back (DB), is a position tasked with covering wide receivers who are running down the field. Their ultimate goal is to make sure that the quarterback has a very difficult time getting the ball to their receiver. This can come in the form of tight coverage, pass deflections, or interceptions. Common names associated with the defensive back or cornerback position include Deion Sanders, Darrelle Revis, Charles Woodson, Champ Bailey, and Jalen Ramsey.
* Safety – the safety (S) can be broken down into two main positions, the strong safety (SS) and free safety (FS). The position is an extension of the secondary, providing extra pass coverage with the cornerbacks. Safeties, while being excellent at covering zones on the field, also have high football IQ and understand the game well, allowing them to play all over the field and help in multiple areas such as run defense, pass defense, coverage, and offensive line penetration. Common names associated with the safety position include Ed Reed, Troy Polamalu, Brian Dawkins, Steve Atwater, Ronnie Lott, and Jamal Adams.

Subsection: Position Abbreviations

* Quarterback - QB
* Running Back - RB
* Half Back - HB
* Full Back - FB
* Wide Receiver - WR
* Tight End - TE
* Offensive Tackle - OT
* Offensive Guard - OG
* Center - C
* Offensive Lineman - OL
* Left Tackle - LT
* Right Tackle - RT
* Left Guard - LG
* Right Guard - RG
* Defensive Lineman - DL
* Defensive End - DE
* Defensive Tackle - DT
* Nose Tackle - NT
* Linebacker - LB
* Middle Linebacker - MLB
* Outside Linebacker - OLB
* Inside Linebacker - ILB
* Cornerback - CB
* Defensive Back - DB
* Safety - S
* Strong Safety - SS
* Free Safety - FS

Data:

Data was collected from multiple different sources in order to develop this project. First, historical data for NFL drafts leading back to the year 2000 were collected. This data was used for training. In order to figure out what players were most likely to be selected at different picks, historical draft data was required. However, because of how much the game of professional football has evolved over the years, this historical draft data was limited to the past twenty-one drafts. The primary reasoning behind this stems from the nature of the game. In recent years, the game of football has evolved to feature primarily pass-heavy offenses on every team in the NFL, making quarterbacks, wide receivers, and offensive linemen very important players. However, there have been spans of time in the NFL where running backs were more valued than quarterbacks or where defensive players were valued more than other positions. With that in mind, draft data was limited to years when professional teams were shifting towards quarterback heavy offenses. These years were arbitrarily chosen by the author.

In addition to historical draft data, the draft order for the 2022 NFL draft, three different player big boards, and three lists of team-by-team needs were also collected. The draft order for 2022 was required in order to predict the results of the draft based on the order of teams making selections. Without this, there is practically no way to predict draft order well.

The three big boards and team-by-team needs datasets are the two most important aspects of this project. During draft season, NFL draft experts, who are primarily made up of former scouts or NFL team executives, put together lists of the best players available who have elected to participate in the draft. All of these lists are based on opinion stemming from player skill, analysis, physical attributes, and an individual’s play on the collegiate football field. The same theory applies to the list of team needs. All teams need to fill some hole in their roster, whether it be a franchise quarterback or a fast wide receiver or a new member of the defensive line. All lists of team needs are also based on expert opinion and can vary depending on who produces the list. Knowing these two kinds of datasets are opinion based, three of each were collected, all before the NFL Combine and Pro Days of athletes entering the draft, and merged together through a similar process. Due to how often these big boards change over the course of time, it made sense to select them all at once and analyze them before the NFL combine and Pro Days. While there may be more movement up and down rankings for each player, relatively speaking all big boards should have similar changes based on performance.

In order to combine the data from different sources, two processes were performed. One was done with the three big boards while the other was performed on the team needs lists.

For team needs, in all three datasets, the tables were constructed similar to the following table, with team name being first and team needs being second, with each position being abbreviated.

[insert table]

However, since there were three lists with varying numbers of team needs listed, they needed to be combined based on priority. It was assumed that items listed early on in the columns were more needed than other positions. With that in mind, points were assigned to the order in which positions appeared. The first position was given a score of ten points, second position a score of eight points, third position a score of six points, fourth position a score of four points, and fifth position a score of two points. Since one of the team needs lists had more than five positions listed out as needs for each team, a cutoff point of five positions needed was made. Although there was a cutoff point, in many instances, more than five positions were listed as needs after the combination of datasets was finished. Below is an example of the final results of the combination of datasets.

[insert table]

For the big boards, data was combined using two main concepts: player ranking and continuity across lists. Each big board contained two hundred players, ranked from one to two hundred. For all three big boards, the player ranked number one was given a point total of two hundred points, the player ranked number two was given one hundred ninety-nine points, and so on. After points were added to each player in each list, the three datasets were checked for the same players. If a player was found in multiple big boards, they had continuity across them. Given their appearance in multiple big boards, players’ points were totaled. For example, if Player X appears in Big Boards A, B, and C with rankings of number one, number three, and number one, that player would have five hundred ninety-eight points (two hundred points for being ranked one, one hundred ninety-eight points for being ranked third). However, if Player Y only appeared in Big Boards A and C, but not B, they would still get their points totaled for the two lists. Lastly, if Player Z only appeared on one Big Board, they would get their points only for that Big Board. This was done primarily in order to keep final rankings of players fair. In theory, a player who appears on multiple big boards is probably well scouted and viewed in a relatively positive manner. Those players should get rewarded for being on multiple big boards. Players who only appear on one big board should still be considered for draft picks, but will be lower on an overall final big board as their point total may be similar to a player listed in three big boards but at the lower end of all three. Theoretically, if a player appeared on Big Board A and had a point total of sixty from only that list, that player would hypothetically hover somewhere around position one hundred eighty on all three big boards. If there was then Player B who appeared on all three big boards around position one hundred eighty (points for position one hundred eighty are equal to twenty-one), this player would have a total of sixty-three points, three more than Player A. One expert’s opinion of Player A may be higher than Player B, but Player B appears more often and more consistently in big boards, giving Player B the same average rating as Player A.

After totaling the points for each player, two hundred eighty-nine players were sorted based on their accumulated point totals. Following this process, point totals were multiplied by position impact multipliers in order to create a second list of player ratings. These multipliers were applied in order to explain positional impact on the field and will be explained further in the following subsection.

Subsection: Positional Impact Multipliers

As stated at the beginning of the Data section, the game of American football has evolved over the course of its history. Unlike the early era of the NFL where running backs were prioritized, the game has shifted to a much more pass heavy offense. With this in mind, along with some other factors, some positions are viewed as more valuable or impactful than others. This is why the Positional Impact Multiplier (PIM) was created. This multiplier is used to show impact based on position in a ranking system. The following tables show both the multiplier for each position and then the rankings of players before and after the multiplier was applied to their total points from the three big boards.

[insert two tables]

As shown in table [number], each position that appeared on the combined big board has a PIM. Positions like the quarterback and offensive tackles get higher multipliers than say a center or running back. This has to do with the aforementioned current playstyle of offenses in the NFL. With quarterbacks being valued at an all time high, as shown through the contract lengths and sizes (in US Dollars) of players such as Patrick Mahomes and Josh Allen, a multiplier of one and a half was applied to quarterbacks ([NFL Active Player Contracts | Spotrac](https://www.spotrac.com/nfl/contracts/)). Offensive tackles have the next highest multiplier as they have the job of protecting the blindside of the quarterbacks. The blindside, or the side of the field the quarterback cannot see due to their back being turned to it, is where often times defensive ends will tackle a quarterback from behind. The reason offensive tackles, specifically left tackles, receive this high multiplier is because they protect the most important player on the team. This impact on the game stems from what Lawrence Taylor, a linebacker for the New York Giants was able to do to Joe Theismann, the quarterback for the team eventually named the Washington Commanders, changed how the NFL viewed both offensive tackles and defensive ends or linebackers (Taylor was viewed as a hybrid player, able to play both defensive end and linebacker). Michael Lewis’s first chapter in his 2009 book, *The Blind Side*, even details the impact Taylor had on the game and how he affected how other teams prepared to play him ([EXCERPT: 'The Blind Side,' by Michael Lewis - ABC News (go.com)](https://abcnews.go.com/2020/BlindSide/excerpt-blind-side-michael-lewis/story?id=9418563)). Due to Taylor’s impact, offensive tackles were now valued even higher than before.

All the other positions have multipliers between one half and one and two tenths. On the lowest end of the spectrum are running backs. With the NFL having shifted away from a predominantly run-based offense, running backs are not as valued as they once were. If a running back does end up getting a lot of opportunities to run the ball, the wear and tear on their body shortens their career. Per Statista, the average career of an NFL running back is a little over two and a half years ([• NFL: average career length | Statista](https://www.statista.com/statistics/240102/average-player-career-length-in-the-national-football-league/)). While this statistic takes into consideration all players who enter the NFL, this average is significantly shorter than a quarterback and lower than the league average of three and a third years. Given this fact, the PIM for running backs is lower than one.

Other positions, such as wide receivers, tight ends, cornerbacks, and safeties, were given slightly higher PIMs because either they contribute largely to the heavy passing offenses of today’s NFL or have the job of providing pass coverage and defense against said offenses. Offensive and defensive linemen have either PIMs of one or slightly higher. Offensive linemen who are not tackles still need to protect the quarterback, making their impact slightly higher than one. Defensive linemen, specifically defensive ends, have the job of getting past those skilled offensive tackles to sack the quarterback, thus warranting a slightly higher PIM.

Methodology:

For this project, three different mock drafts were created and the descriptions of those drafts will be provided in subsequent subsections. These mock drafts, the Machine Learning Mock Draft, Team Needs Mock Draft, and Team Needs Plus Mock Draft, all aim to predict draft selections as accurately as possible. However, it is important to note that these drafts, specifically the Machine Learning Mock Draft, will most likely end up being inaccurate for reasons that will be identified in each subsection. Additionally, for each draft, because of when the data for this project was collected, there are numerous factors that could change for each mock draft. Those factors are trades of either players and/or draft picks, free agency signings, unforeseen trades on draft night, and prospects whose rankings by experts went up or down after the NFL Combine.

Subsection: Machine Learning Mock Draft

For the first mock draft, historical data was used. This mock draft, named the Machine Learning Mock Draft (MLMD), was built by using the historical data for each pick position in the past twenty-one drafts to find the percentage chance of a player type being selected at a certain pick position. Below is a theoretical table of what the historical data could produce.

[insert table]

In the example, [percent] percent of the time the first pick in the draft is a quarterback. With that in mind, we know that more likely than not, the team with the first pick in the draft will select what is deemed by that team’s front office and scouts as the best quarterback in the draft. However, in the case of the 2022 NFL Draft, we run into an issue here. Jacksonville, the team with the first pick in the 2022 NFL Draft, selected a quarterback in 2021 with the first overall pick [ref]. Because Jacksonville selected Trevor Lawrence, a quarterback, in 2021, it is very unlikely that they would select another quarterback in the following draft. In fact, in the past twenty-one drafts, the only instance of a team drafting a quarterback in the first round in back-to-back drafts happened in 2018 and 2019 when the Arizona Cardinals drafted Josh Rosen and Kyler Murry, respectively. With this in mind, there is already a problem with using a mock draft built entirely on historical data and machine learning, thus there becomes a need for a different method of making a mock draft. While there is a chance of teams selecting a player with the same position as one they drafted the year prior, this machine learning draft completely ignores any team needs that exist.

Subsection: Team Needs Mock Draft

For the second mock draft, it was apparent that strictly using machine learning would not produce a fantastic result. Team needs play much more of a role in draft picks. While it would be possible to build a model to determine what needs a team has, that was not the goal of this project. In order to work around this need to include team needs, the Team Needs Mock Draft (TNMD) was built as a follow-up to the MLMD. In the TNMD, the combined list of team needs and big boards were used in order to determine what picks a team might make. In Table 1, we can see that Jacksonville has offensive tackle (OT) listed as their top team need. Knowing this, it would make sense for them to draft the best offensive tackle available Jacksonville in this situation would most likely select Evan Neal, an offensive tackle from the University of Alabama. However, this mock draft does not include the PIM metric. It strictly looks at the best player available at the first need listed for a team. Jacksonville’s second team need is an offensive guard and its third need is a wide receiver. If there was a wide receiver or offensive guard with a higher player score than the best offensive tackle because of the PIM multiplier, Jacksonville may select the wide receiver or offensive guard because that player is more valuable to the team. With this in mind, a third mock draft was built.

Subsection: Team Needs Plus Mock Draft

For the third mock draft, the PIM metric was needed in order to properly account for the impact players could have on the field. Using Jacksonville again as the example, we know they need an offensive tackle, offensive guard, and wide receiver. If we include PIM in player rankings, we can see that quarterback Kenny Pickett from the University of Pittsburgh rises from twenty-first place to first place. We can see some wide receivers like Drake London from the University of Southern California or Chris Olave from Ohio State University fall down the rankings board a handful of spots. Is this system of including the PIM perfect? Not at all. However, it does account for the impact a player may have. With this system in mind, Jacksonville most likely would still select Evan Neal, however, teams with picks further down in the draft may end up considering different players with higher PIM scores who are not their number one team need.

Results:

Analysis:

References: