Machine Learning in the NFL: Predicting Future Football Stars

Aneesh Sunkarapalli
Virginia Tech
aneeshsunk03@vt.edu

Rudra Patel
Virginia Tech
rudy@vt.edu

Figure 1: Most recent NFL Draft, hosted in Detroit, 2024.

ABSTRACT

In the field of sports analytics, the NFL Combine presents a unique environment of data; it possesses the potential to highlight different trends and common statistics among players at the highest-tier. Using this data, we examined the factors that lead to success in the athletic world. We utilized various machine learning techniques to identify the features of combine data that most significantly influence an athlete's likelihood of being drafted into the NFL.

KEYWORDS

datasets, classification, predictions, sports, statistics, machine learning models

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

ACM Reference Format:
Aneesh Sunkarapalli and Rudra Patel. 2018. Machine Learning in the NFL: Predicting Future Football Stars. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

1.1 Motivation/Context

The National Football League (NFL) Combine is an annual, significant event where college football players are evaluated on their skillset, both physically and mentally, through a series of robust drills. The Combine is an event that has a large influence on their likelihood of being drafted with its status being boiled down to a strictly-structured hunt for prospective players at the national level (the NFL). The players that are invited to essentially "perform" at these events are already known to perform well in a real-game scenario; scouts and recorded game data are all considered when it comes to determining which players should be invited to the official NFL combine. This idea that only the best-of-the best are invited says a lot about not only the prestige of the ceremony/event itself, but of the players too. These players are all known to have some sort of promise when it comes to having a future in professional football, far extending their college residence.

Still, just being really good at the sport won't cut it. The roster from team-to-team on the NFL can't just recruit every player they find to be skilled. There needs to be a certain threshold or level of play exhibited by the set of players to make them be considered for the draft, and therefore a placement on a team to play football at the national level.

1.2 Description of the Problem

The overarching problems stem from the context: with the annual draft not taken place yet, can we figure out which of the most recent set of players that attended the combine will be drafted? How can we determine this? And how would we explain it?

As we broke it down, we started looking at different, pre-existing groups within the data. The most prominent groupings we found in the context of our project were the Power 5 conference school players and the Non-Power 5 conference school players. For context, the Power 5 conferences consist of the Southeastern Conference (SEC), the Atlantic Coast Conference (ACC), Big Ten Conference, Big 12 Conference, and Pac-12 Conference. Each of those conferences represent a series of different schools, which all have their unique player rosters for the sport of football. The Power 5, as you could assume, is the most prominent 5 conferences out of all the possible ones in the sport. This is why we're considering them as a group, and placing the remainder of players from other schools (Non-Power 5 conferences) into their own category. This could help lead to different ideas for solutions to our problem and help us formulate an explanation to certain findings we have in terms of who's drafted and who's not, and if the prestige of the school, therefore the conference, that they're a part of has some sort of correlation with their draft status.

1.3 Initial Obstacles and Ideas for Solutions

The obstacles we had in constructing ideas for solutions to the aforementioned problems pertained to how we would acquire the data, normalize it, and feed it into our normalization for analysis beyond basic data analysis methods.

Our primary ideas for ways we could approach the problem and implement solutions revolved around utilizing various classification machine learning models and visual-based machine learning models to satisfy the objective of identifying a binary output of whether a player is drafted or not. Multiple types of classification models would allow us to check ourselves in terms of cross-referencing the results of the models to see if there's any noticeable disparities in the results, and if/where they're stemming from. The visual-based machine learning models will provide some information to build explanations for our findings, but one of their main uses is to allow us to actually see and process the answers to our questions at hand; having a visual output that clearly identifies groups of the data makes it appealing and easier to understand what the model and data is saying in the context of our project.

2 RELATED WORK

2.1 Similar Projects

Although our project is focused on a different topic, we found various similar projects online such as predicting NFL success from the Northwestern Sports Analytics Group, articles and work from Google Scholar related to predicting the results of NHL games, and even monitoring public sentiment of NFL draft picks using machine learning.

The Northwestern Sports Analytics Group also utilized machine learning to specifically look at the current year's (2024) college quarterbacks and receivers, only 2 positions, and what their success might look like in terms of getting drafted into the NFL [1]. The study employed various regression models to determine features that can predict success, measured as the chance of being initiated into the NFL via the draft process, and came to different conclusions revolving around their outcomes for each player position. They compared the existing draft order with their outputs from the model and found that the quarterbacks they identified to be the "most successful" in the 2024 NFL Draft seemed to line up with the exception of an outlying player.

The Google Scholar article about NHL game data analyzation highlighted how one group utilized big data and machine learning in tandem to look at web-scraped player-team data in order to create a predictor for game outcomes [2]. They attempted to try to produce an output looking at the possible scores given a certain roster of players and the team, which eventually seemed to expand towards looking at individual player statistics for the best possible recruitments across NHL teams. A deeper layer of this project was looking at what the salaries of members of these teams should be altered to, if it all, as well. By exercising different forms of analysis, such as PCA and nonparametric statistics, on top of a support vector machine and ensemble ML algorithm, the authors built their predictor system. The support vector machine reached an accuracy of above 90% with the inclusion of the ensemble algorithms they used, indicating great success in their goal of predicting which team would win a certain game based on certain parameters.

The public sentiment monitoring of NFL draft picks using ML was conducted by a single member of Elizabethtown College named Jonathan Wiseman who attempted to analyze general website users speaking about the NFL draft, both positively and negatively, to apply it to a machine learning model [3]. Its primary use was to process different predictions and reactionary information from the public to generate useful information about various players and their drafting status. The amount of data is abundant; so many members of the public take to online locations such as social media to rant about their thoughts and opinions on how a plethora of aspects of the sport are currently going, which includes players that are up to be drafted, general thoughts on teams, and even coaches themselves. By feeding this data into his model, Wiseman created something that takes the public's word and quite literally puts it to test with the real statistics or outcomes of different games and/or events (drafts). Some of these connections include looking at how people spoke about a player's performance versus their overall in-game performance, as well as the general feeling on players prior to their performance in an imminent game.

2.2 Differences and Limitations of the related works

Other similar sources we found online were more informal. These projects were on blogs, personal websites, and GitHub projects, and weren't thoroughly consulted by us when deciding how to approach our project as a result; they were promising, but weren't long enough or at the level of complexity we felt like this problem would require.

Looking at blog post from author Brock Grassy, the project outlined a similar objective to ours [4]. He used web-scraping as well to accumulate all of his data. His outcomes using XGBoost, Random Forest, and various linear regression models proved to be successful for the constraints of the problem; as Grassy stated, "The model's top 32 players included 15 of the first 32 actual draft picks, as well as 17 of the first 34. However, the overall "report" aspect of this project was not quite as extensive as the data and actual machine learning process implemented. Particularly, the most recent class of college football players entering the draft season were not considered/scraped as part of the dataset so they could not be discussed when it came to the conclusion of the project. This is one limitation of this project in compared to the approach we hope to take.

Another smaller-scale project was found on a GitHub repository created by Sean J. Taylor [5]. He also followed suite in web-scraping from a football player statistic database, then used it to create models that would help him predict the percentage that a player will be a first round draft pick. While the amount of data points available for the training process was large, there was an immediate issue of what to do about the players that are included within the data but don't attend an NFL combine, therefore having missing values that need to be filled in for the ML process. While those values were imputed, it leads to inherent flaws in the design of the computational aspect of the project. The results for this project were also successful/provided solid accuracies, but its use case is where it's limited; knowing the probability of a player being a first-round draft pick is only useful to a certain extent. Still, this project was good work overall and the author even indicates that if this project were properly expanded using different models for valid predictions, it would be much better.

We wanted to model our strategy of approach in a similar fashion to the previously mentioned related works. However, they each possess their limitations which lead to our final approach and project being marginally different in their qualities. For example, while the Northwest Sports Analytics Group report was well done, an aspect of their project that was a limiting factor is making their sole focus players in the quarterback position, excluding every player in any other position. The NHL project defined by the Google Scholar article could've utilized more features/metrics as their basis of criteria, as well as included more team performances for both. The public sentiment monitoring project conducted by Jonathan Wiseman is effective, but if it incorporated other models and player statistic data like we hope to, both outputs can be used in collaboration with one another to create a machine that has more powerful predictive capabilities.

3 METHOD OF APPROACH

3.1 Current Approach

Taking in all the factors from the description of the problem and the obstacles it posed, our ideas for solutions, and the related works' strengths and limitations taken into account, our idea is to do some analysis to find basic findings in our data to determine its significant features and nuance, then utilize an array of classification and visual-based machine learning models to look at different layers of being drafted, like what metrics lead to it (combine drill features and school/conference prestige) and if it actually occurs or does not occur.

3.2 Data Processing

Our dataset was obtained from Kaggle and contained the NFL combine stats and general stats (height, weight, school, etc.) of college football players and included a binary classification to indicate as to whether they were actually drafted into the NFL or not [6]. To expand upon this and make it more modernized as our approach indicated, we web-scraped a premier football statistic website to append to the data the most recent (2024) college football player NFL combine and general statistics to later use our models to generate draft predictions for those players, although the official draft hadn't occurred at the time of inception of this project so we couldn't look at our output vs. the truth.

3.3 Basic Data Findings

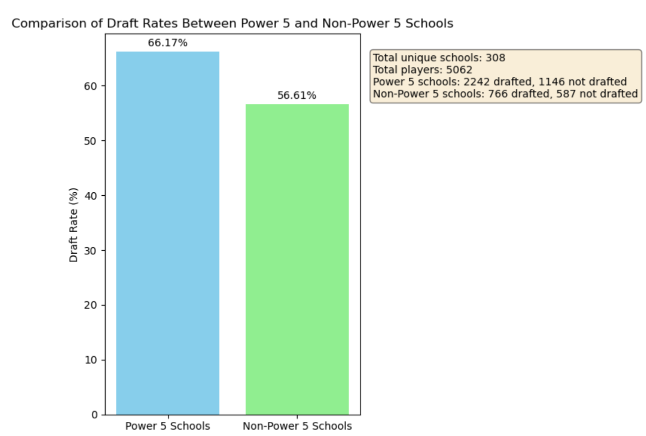

Figure 2: Power 5 vs. Non-Power 5 Draft Rate

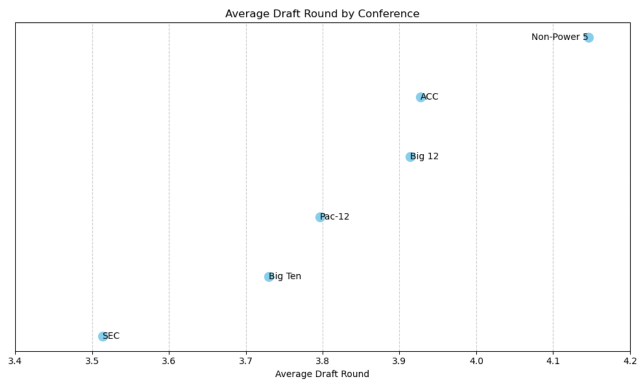

Figure 3: Average Draft Round by Conference

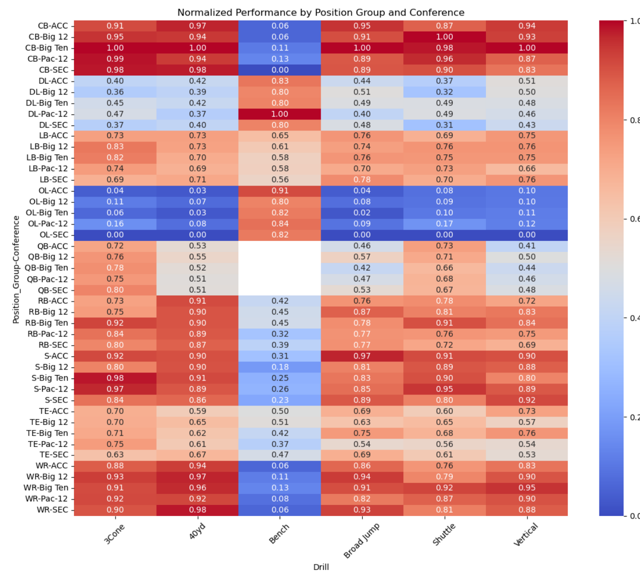

Figure 4: Player Position Statistic Heatmap

Our analysis revealed notable differences in the our data for NFL draft outcomes. Key findings, for players from Power 5 and Non-Power 5 conference schools, were that Power 5 school players had a significantly higher draft rate (66.17%) compared to their Non-Power 5 counterparts (56.61%). This is clearly evident via the bar graph visualization, which distinctly shows the draft rate disparity and suggests a potential advantage for Power 5 conference school players.

Across the six evaluated groups, SEC players were drafted earlier (average draft round 3.51) than the Non-Power 5 average (4.15). This prompted further examination, providing insights into the average draft rounds per conference within the Power 5: The SEC's average draft round was markedly lower than other conferences, indicating earlier drafting of their players. Along with this, the SEC emerged as a clear outlier, prompting more investigation into the conference-specific dynamics. Following this analysis, we were encouraged by the SEC's outlier status to further compare the Power 5 conferences against each other. Our results underscore the potential advantages of being in a 'superior' conference in the realm of competitive football. This basic, intial data analysis suggests that conference affiliation may play a significant role in a player's journey to the NFL.

Our investigation into whether affiliation with a specific Power 5 conference influences combine drill performances across different player positions yielded the following insights: Players from the SEC demonstrated exceptional performance across various combine drills, aligning with the trend of SEC players being drafted earlier on average compared to other groups. Despite this, the SEC's dominance was less pronounced in this aspect compared to other areas we explored. The heatmap analysis revealed significant variations in performance both between conferences and within specific positions. For instance, SEC's wide receivers and tight ends showed superior speed, while defensive linemen from the PAC-12 excelled in the bench press. The data highlights that each Power 5 conference exhibits distinct strengths and weaknesses across different drills and player positions. No single conference showed overwhelming superiority across all areas. Although a potential correlation between a conference's average draft round and its players' drill performance was observed, conclusively proving this relationship requires further detailed analysis. Our findings illustrate the competitive landscape within the Power 5, where each conference showcases its unique talents and competencies in various skill tests, contributing to their players' prospects in the NFL. In essence, while specific patterns and strengths emerge, the diversity and balance among the Power 5 conferences underscore the complexity of drawing definitive conclusions about the impact of conference affiliation on combine performance and NFL draft outcomes as we initially hoped to.

3.4 Machine Learning Model Evaluation

As previously highlighted, an integral part of our approach to the central problem is to create visual representations to make it more accessible in terms of understanding what the data is saying after feeding it into different types of machine learning models. Another aspect of how we would go about our actual experimental evaluation is if we could have some sort of connection between any of the models to cross-reference or verify their results. With all these considerations in place, along with our current method of approach and the outcomes of the data analysis/what it told us about the features of our data, we arrived at the following four ML models for evaluation: a logistic regression model to display feature importance and potential drafting outcome, a random forest model to display feature importance and drafting outcome, a K-Means Clustering model to display player statistics and draft rate similarity, and a decision tree to display draft statuses based on combine drill statistics.

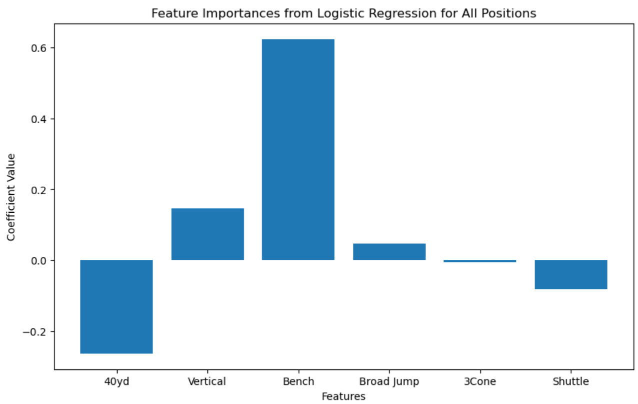

Figure 5: Logistic Regression Feature Importance plot

4 EXPERIMENTAL EVALUATION

4.1 Logistic Regression for Feature Importance and Potential Drafting Outcome:

We first fed our data into a logistic regression model to determine feature importance (drill significance) for successfully getting drafted into the NFL.

```
Predictions for 2024 players:
      Year              Player  Predicted_Drafted
4741  2024  Kris Abrams-Draine                  1
4742  2024        Isaiah Adams                  0
4743  2024         Rasheen Ali                  1
4744  2024           Erick All                  1
4745  2024        Braelon Allen                 1
...    ...                 ...                ...
5057  2024        Roman Wilson                  1
5058  2024         Mekhi Wingo                  1
5059  2024       Xavier Worthy                  1
5060  2024       Jaylen Wright                  1
5061  2024          Zak Zinter                  1

[321 rows x 3 columns]
```

Figure 6: Logistic Regression Draft Prediction Output

Following this, we exercised the model to produce a binary output for 2024 draft predictions, where 1 indicates they're predicted to be drafted and 0 indicates they will not be drafted. This output encapsulated the current class of college football players that exhibited their skills at the combine.

The model demonstrated an overall accuracy of 69%, successfully predicting draft outcomes for 69% of the players. It showed high precision in identifying players who were not drafted, with a precision rate of 72%, although its recall for this group was low at 21%, indicating that many undrafted players were not accurately captured by the model. Conversely, the precision for predicting drafted players was reasonable at 69%, and the recall was impressively high at 95%, suggesting that the model effectively recognized most players who were drafted.

In terms of feature importance, the Bench press had the most positive impact on the prediction of draft outcomes. Conversely, the 40-yard dash and Shuttle runs were associated with negative coefficients, though these results require careful interpretation. The Vertical jump, Broad jump, and 3-Cone Drill had a marginal positive influence on the predictions. It's important to note that the model applied a uniform approach to all player positions; incorporating position-based filtering might enhance the predictive accuracy, particularly for specific positions like linemen, which may exhibit unique patterns of feature importance.

4.2 Random Forest for Feature Importance and Potential Drafting Outcome:

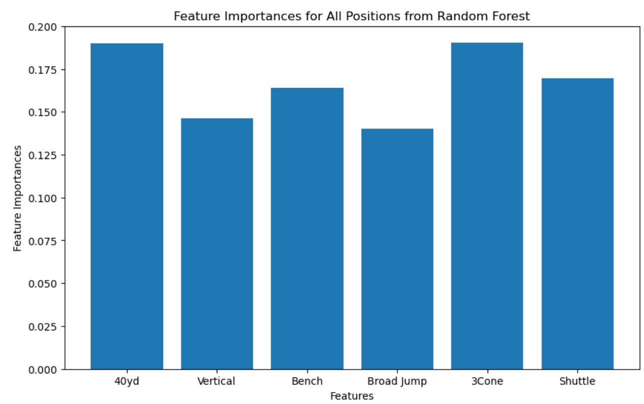

Figure 7: Random Forest Feature Importance plot

The next part of our approach was using Random Forest for the same reason that we used logistic regression: determine feature importance and potential drafting outcome for players. This was important to bring into play because it gave us two ways to produce outcomes for the same problem, allowing us to see if there's any experimental error when it comes to our evaluation.

The model exhibited a modest overall accuracy of 68%, which is slightly lower than that achieved by Logistic Regression. For predicting non-drafted players, it demonstrated moderate precision at 58%, indicating the presence of some false positives, while the recall improved to 34%, marking a better ability to identify non-drafted players compared to Logistic Regression. Precision for drafted players was comparable to Logistic Regression at 70%, with a recall of 86%, showing that most drafted players were correctly predicted, albeit at a lower rate than Logistic Regression.

In analyzing feature importance, the 40-yard dash emerged as the most influential factor in predicting draft outcomes. Other features like the Vertical jump, Bench press, and Shuttle runs were also important but to a lesser extent than the 40-yard dash. The 3-Cone Drill was identified as the least important feature within this model.

4.3 K-Means Clustering for Player Stats and Draft Rate Similarity:

We applied K-Means clustering to our NFL Combine data from 2010-2023 to group players into five distinct clusters based on their

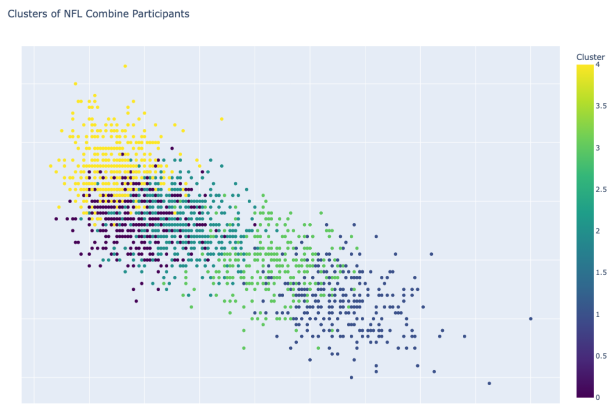

Figure 8: (Interactive) K-Means Clustering plot

performance metrics (feature selection: 40yd, Vertical, Bench, Broad Jump, 3Cone, Shuttle) The K-Means algorithm helped group players with similar combine performance profiles, and the clusters show their respective likelihoods of being drafted.

Our clustering output is interactive and provides key summary statistics for each player when you hover over each point in our code. This allows users to have a closer look into how each player might be separated into their own designated cluster out of the 4 in total with draft rates of 0.5344827586206896, 0.5616438356164384, 0.6538461538461539, 0.7151515151515152, 0.7723214285714286 respectively.

4.4 Decision Tree of Draft Statuses based on Combine Drill Stats:

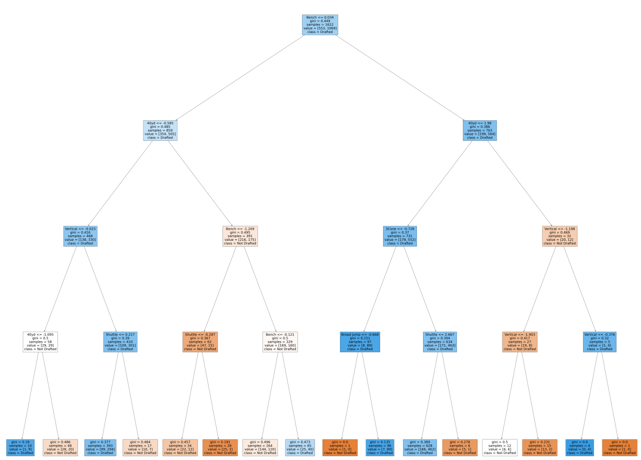

Figure 9: Decision Tree based on feature (combine drill) threshold

We utilized the decision tree model with the NFL Combine data from 2010-2023 to group players into five distinct clusters based on their performance metrics (same features of: 40yd, Vertical, Bench, Broad Jump, 3Cone, Shuttle) Much like other decision trees, each node pivots around a feature's computed threshold (calculated in the data training process). In our case, the splitting criteria was based on each drill statistic. The goal of the decision tree was to determine draft status based on the class of each node; as the depth of the tree increased, so did the specialization of players that would fall into different sections of the tree based on their respective performance in the drills (our features). That is, if a player performed in less than or equal to the node's threshold, they would fall into the left and those over to the right as per the standard. The nodes would be classified as drafted or not drafted, ultimately providing a visual representation of the ratio of players drafted or not drafted into the NFL once the last level of the tree is reached. Since the max depth of the tree was initialized at a level of 4, the visual representation of the tree (output) is quite expansive.

5 CONCLUSION

We found that Power 5 school players demonstrate higher draft rates and are selected earlier in the draft, especially from the SEC. Performance data indicates strengths vary across conferences and positions, and the competitive nature of Power 5 conferences contributes meaningfully to players' NFL draft prospects. Players from Power 5 conferences exhibit a higher likelihood of being drafted and are drafted earlier compared to their Non-Power 5 counterparts. This suggests that the prestige and resources associated with Power 5 schools may play a significant role in enhancing draft prospects. Furthermore, since the analysis of physical abilities shows that there isn't really a strong impact on whether or not an athlete is drafted, refuting arguments that Players from Power 5 conferences are being drafted because they are more physically capable.

Despite the SEC's notable success in achieving the earliest average draft picks, our comprehensive drill performance analysis via heatmap visualizations underscores that each conference, including those within the Power 5, showcases unique strengths and weaknesses.

The data reveals a competitive equilibrium among the Power 5 conferences, challenging the notion of a single dominant conference. This competitive landscape underscores the collective excellence of Power 5 schools in preparing athletes for the NFL, as evidenced by superior draft rates and combine performances.

Our data analysis showed some interesting takeaways about the importance of things like athleticism and combine performances. For one, our data analysis suggested athleticism and performance in drills were not the biggest factor in a player getting drafted, supporting the notion that other traits such as player-team fit, mental attributes, and of course the school they played at, can all have significant effects as well. This helps give context to long-accepted beliefs about certain schools or conferences providing different levels of readiness for the NFL, while also confirming that these notions are just a part of what goes into drafting players.

The draft process is very holistic, meaning a variety of factors show different impacts as measured in our analyses, but we can conclude that some of these are qualitative factors, such as school/conference prestige, culture fit, personality and intelligence gauged in draft interviews, etc. These factors will not be represented in our data the same way that metrics on players' speed, agility, and strength can, but even without knowing exactly what these factors are we can see their impact in the data and infer on the nature of their impact.

This all tackles the nuance found within our data and the analysis we did upon it. Now, we need to delve into what our machine learning process showed and affirmed to us in terms of our initial goal and what we found based on our produced models.

After applying various ML techniques, our understanding of what influences NFL draft outcomes has expanded greatly. Our logistic regression model revealed primary predictors of draft success, lining up with our earlier discussions about the impact of combine performances. For example, it affirmed that metrics like the bench press have the effect of swaying drafting decisions.

The random forest model introduced another layer of insight by supporting the statement that the 40-yard dash is a significant predictor as week. The difference in model outputs is an indicator of the variations in nature of athlete stat evaluation; both speed and strength may indicate potential success differently from model to model.

The decision tree and K-Means clustering models exemplified the drafting landscape by classifying players into distinct groups based on performance and demonstrating how these placements correlate with draft outcomes. In particular, the decision tree provided a visualization of the decision-making thresholds based on combine drills, providing an idea about the logical aspect/blueprint that's naturally behind the drafting process.

The ML explorations supported our approach and provided a unique outlook on the drafting process. These insights have also highlighted the potential of integrating more analytics into the current standard methods of scouting, which is another way players are examined on their skill and ability for drafting.

In terms of improvements, providing our models with additional layers of data such as player injury histories, overall college performance stats, and some sort of mental ability assessments could add more levels of insight to our project and give our predictions a greater weight.

6 FUTURE WORK

While our models and results proved to be relatively successful in analyzing features and predicting draft rates for different college football players based on their performance across various combine drills, and potentially their school's conference, there's still a plethora of additional machine learning models that can be effectively applied to our problem to produce more concrete results.

Potential future uses of these models for this problem and the approaches we implemented have a diverse range; however, the two primary uses we concluded that this project could lead to is the incorporation of natural language processing (NLP) and neural networks.

NLP can be used to investigate sentiment analysis surrounding draft picks, and the NFL in general, on online platforms such as news articles, social media, and commentary. The response on these platforms could help gauge public and expert opinions to identify undervalued talent and untapped potential from different players on the field in the college football scene. A powerful enough model powered by natural language processing could have a high chance of successfully filtering input from valid/verified accounts and news sources in tandem with existing models we implemented. This process would essentially be consistent data-scraping of the web, which would amount to a large amount of data, but at the risk of finding "outliers" in terms of incorrect sentiment being considered. This could look like accounts on Twitter or other forms of social media saying objectively false information for the purpose of messing around or out of anger, which would skew what the model's computations are. The actual application of this in the context of our problem could essentially assign a draft rate percentage for different college athletes, computed through other methods, and then use their NLP model to dynamically update this percentage throughout the season based on different events occuring in which said players can exhibit their skills, such as official games, skirmishes, or additional, small-scale combines. The process would conclude immediately prior to the next annual NFL draft.

Neural Networks can be used to analyze movement patterns from video data, such as archive game footage and documented practice film. Both of these forms of video data are already recorded so accumulating this content wouldn't be a significant cost in terms of expending resources to produce a functioning neural network. Neural Networks could help identify and evaluate player techniques, agility, and decision-making in real-time game scenarios to ultimately predict player performance before games even finish. This capability is powerful; it expands upon the initial problem of just assigning a draft status and possibility percentage and can be used to determine how certain players might perform in terms of scoring in different categories of the game. Looking at different potential thresholds of scoring metrics, such as touchdowns and running yards, these can be extended into the world of sports betting. Predictions produced by a functioning neural network could let users know what projected outcomes of games are based on an abundance of factors, such as the combine data that we used and in-season game performance. A feature of the neural network could be comparisons with historically successful players and their movement patterns to produce factors in draft chance and classification predictions. Overall, an efficient/resource-conservative approach to implementing a neural network for this problem would be ideal, but it might require more assets to reach its promise.

REFERENCES

[1] Posted By: bag2351. (2024, March 29). *Using machine learning and college profiles to predict NFL success.* Northwestern Sports Analytics Group. Available at: https://sites.northwestern.edu/nusportsanalytics/2024/03/29/using-machine-learning-and-college-profiles-to-predict-nfl-success/

[2] Gu, W., Foster, K., Shang, J., & Wei, L. (2019). A game-predicting expert system using big data and machine learning. *Expert Systems With Applications, 130*, 293–305. https://doi.org/10.1016/j.eswa.2019.04.025

[3] Wiseman, J. n.d. *Monitoring public sentiment of NFL draft picks via machine learning techniques.* JayScholar. Available at: https://jayscholar.etown.edu/comscistu/1/

[4] Anonymous. n.d. *Applying machine learning to predict the NFL draft.* Available at: https://www.brockgrassy.com/blog/applying-machine-learning-to-predict-the-nfl-draft/

[5] Taylor, S. J. n.d. *Learning the NFL draft.* Available at: https://seanjtaylor.github.io/learning-the-draft/

[6] NFL Combine - 2010 to 2023 performance data. (2023, December 15). *Kaggle.* Available at: https://www.kaggle.com/datasets/thomassshaw/nfl-combine-performance-dataset