Determining Importance of Combine Performance in NFL Draft

Joshua Byrne, Cristobal Garcia, Drew Hasse

1 Abstract

In our project we are attempting to create a classifier that is able to predict whether or not a player will be drafted solely based upon their performance at the NFL combine. Every year there is an extremely large amount of interest surrounding the combine, with both fans and teams eager to determine when or if a player should be drafted. The statistics we used were mostly related to physical ability, such as lifting, running, and agility. The amount of money and research that goes into player performance is continually growing, which is why we hoped to determine the overall importance of these factors on a player's prospects in the NFL.

Our initial approach was to use decision trees to develop a set of rules that would classify players based upon their physical abilities as well as position, college, height, and several other features. We believed this, or nearest neighbor, would allow us to create accurate models based on the basic intuition that the physical ability was highly correlated with draft prospects. This, however, proved to not be the case as we found very little improvement with these models over ZeroR. From here we began to spend a great deal of time on transforming our dataset to interact variables and spread out possible distributions. We found this to improve the accuracies of several classifiers we explored, but by very little.

Even through the use of decision trees, logistic regressions, and many other models, our best classifier only proved to be 2 % more accurate than ZeroR. We discuss these results in more details, as well as their several implications in the attached report.

2 Dataset

We obtained the original dataset from pro-football-reference.com. It contained NFL draft combine statistics for all 3618 combine participants from 2008 to 2018. The data for each player included player position, height, weight, school, specific workout statistics, and the round they were drafted in (if drafted at all). After we obtained csv files for each combine year, we wrote several python scripts to make them compatible with Weka (removing apostrophes from names, changing 2nd to 2, etc.) We also wrote a script to normalize all numerical statistics for each individual combine between 0 and 1. We chose to do this because athletes' performance on combine workouts has improved over time, thus failing to normalize the data would prevent our machine learning models from achieving high accuracies over a large number of years. Finally, we randomly split the data into a training set with 75% of all data (2712 examples) and a holdout set with the remaining 904 examples to be used for the final evaluation of our models.

3 Initial Approach

Using the intuition that physical ability is the primary factor in if a player is drafted or not, we believed that decision trees would be able to create a clear and concise set of rules that would allow for a highly accurate classifier. Using the normalized statistics we believed any bias of time or a particularly athletic year would not cause our classifier to be inaccurate. As we began to run our models in Weka, however, it clearly became apparent that our mostly numerical features were not as informative as our intuition led us to believe.

Our preliminary runs of the data using the J48 classifier (decision tree with pruning) returned a tree that simply matched the ZeroR classifier, implying that our data was almost entirely uninformative. The accuracy for both ZeroR as well as for the J48 model was found to be 62.6613%. In order to gain some information from the J48 model, we proceeded by removing pruning from the model so that we would be able to observe the information gain provided by the variables, even though our accuracy was reduced. This revealed that position was the most informative feature within our dataset. We agreed with this discovery, as the relevance of the performance statistics is highly dependent upon the position of interest. For example, it is far more critical that a lineman have a large bench press than it is for a quarterback. This was the basis for our future exploration that is discussed in the Data Adjustments section below.

4 Data Adjustments

At this point in time we additionally began to take the step of simplifying our model by no longer predicting the specific round a player would be chosen in, but rather if a player would be drafted at all. This simplification increased the accuracy of our classifiers, but also of ZeroR. With this simplified classification task, we ran several other models to simply see if they were better able to classify our data.

Before transforming the data J48 achieved 37.3387 % accuracy on the training data (10-fold CV). After switching to a binary class, J48 achieved an accuracy 62.6613 %.

SimpleLogistic was able to increase our 10-fold cross validation classification accuracy to 64.4567 %. Another model that we tried on the binary data was a Bayes Net, but this model only achieved a worse accuracy than our perceptron model (63.9882 %). This did not seem like a high enough accuracy for predicting whether a player was drafted or not, so further data adjustments were pursued.

Moving forward we decided to look at Z-scores rather than the combine numbers themselves, and we also added attributes for each combination of position and statistic. Originally position was an attribute and there was one attribute for each of the combine workouts. After our data adjustment we had attributes like WR-Bench, QB-Bench, LB-Bench, etc. Then each player would only have the attributes for their position filled in. This adjustment was looked into because we wanted our model to be able to capture the differences in statistics across positions for all models we were testing. For decision trees, this is unnecessary because the tree will split on position first, but for running something such as logistic regression, we thought this step might be useful for exposing these differences to the model. Along with this transformation, we calculated the Z-scores for each statistic and used those instead of the raw numbers. The idea behind this was that the Z-scores (generated using another python script we created) would be useful in identifying players who vastly outperform others since high Z-scores indicate the player is performing much better than the mean. Conversely, a Z-score would also reveal the players that performed poorly.

Despite implementing these data adjustments, the maximum accuracy we could find did not improve much. Most models we ran were achieving accuracies of around 60-65%. This may indicate that our task is not highly predictable due to other factors influencing draft outcomes.

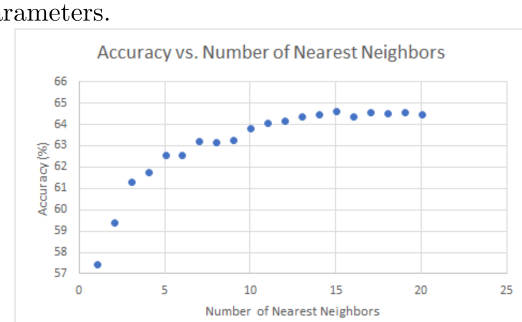
5 Final Result

Table 1 displays several classifiers built using our final transformed training data and tested on our withheld dataset. The results show that our transformation of the features did allow some of the probabilistic models, such as Naive Bayes, to

Classifier	Test Set Accuracy (Withheld Data)
Zero R	61.0619 %
J48 (Decision Tree)	61.0619 %
J48 (w/o pruning)	62.2788 %
Naive Bayes	63.7168 %
Bayes Net	62.2788 %
Logistic	52.6549 %
10-NN	61.1726 %
Random Tree	60.9513 %
Random Forest	61.0619 %
REP Tree	59.6239 %
Stochastic Gradient	61.0619 %
LogitBoost	62.8319 %

Table 1: Result of Classifiers on Withheld Test Data

classify the data slightly better than ZeroR. That said, some of the classifiers which we had previously evaluated using cross validation showed a reduced classification accuracy, further limiting the improvements we had hoped to see after working to reshape our entire dataset. The figure below further shows our difficulties in improving the accuracy of our classifiers through the example of nearest neighbor. Even with very large feature space and training set we were barely able to improve overall accuracy as we adjusted model parameters.



The results of our classifiers seem to reveal that performance at the NFL combine is not a key factor in determining if a player will be drafted or not. From this, one could conclude that the NFL combine does not provide relevant insight into a player, and therefore should be forgone. We believe the issues with our classifiers to be more nuanced, however. It is possible that being seen at the combine is just as important as specific performance, and it is used solely to reinforce already held opinions of the players based upon their college performance. Additionally, some players participate in individual pro days at their own schools, which are very similar to the combine in testing a player's physical abilities. We believe that the plethora of information available about each player prior to the combine renders the specific combine performance irrelevant, and simply turns it into a spectacle at which a player can make themselves be seen by millions. Our findings do not prove the combine to be irrelevant, but do show that the specific physical performances are not the key spectacle that they are made out to be.

6 Future Work

For future work, if our goal was to build a more accurate model, we could include college statistics in our dataset. The importance of these statistics would further differentiate draft-worthy players from those who are not. The college statistics might even be informative enough for models to accurately predict draft round. That said, this is almost a separate classification task than determining the importance of combine performance.

7 Individual Contributions

Each member of the group contributed to each step in the projects progress. All of us were able to contribute to each of the sections discussed above. If some sort of delineation of work is required, it could be said that Cris lead the data collection phase, Josh took the lead on the data transformation, and Drew took initiative in running the models.

8 References

Sports Reference LLC. Pro-Football-Reference.com - Pro Football Statistics and History. https://www.pro-football-reference.com/. (2018)