Dylan Minarik, Ryan Rose, Steven Vorbrich
EECS 349: Machine Learning
Northwestern University

NFL Playoff Predictor

Motivation

Our team designed an algorithm to determine which NFL teams would make the playoffs using statistics and results from previous years. In the last year, over $1 billion was spent betting on football games in Nevada casinos alone, and an estimated 33 million people participated in fantasy football leagues. We were curious if there were certain statistics to weigh more heavily when attempting to predict which teams will have a good season. The widespread implication of this project is thus evident, with the large percent of the US population with vested interests in the NFL.

Dataset and Manipulation

The team statistics we used for our project came from *pro-football-reference.com*, a website containing a multitude of statistics regarding the NFL. We decided to download csv's containing a number of offensive and defensive statistics for each NFL team since 1998, for an initial dataset of 576 lines with 52 total attributes (*Appendix A*), and an additional binary attributes added, stating if the team made the playoffs in the next year, which we're trying to predict. The data from 2013 and 2014 (64 instances in total) were set aside to be used as the test set for our experimentation, and the data from 2015 were set aside (as they don't have any playoff predictions for the upcoming year).

Some initial decisions were made by our team to cut irrelevant attributes, such as number of games played (16 for everyone), and the ones regarding penalties and penalty yards. Once the pre-processing was completed, we were left with 42 attributes in our dataset, which is the set we initially decided to use for training and testing. However, we were aware that with a large number of attributes such as this, our results would be very susceptible to overfitting, which we kept in mind moving forward with testing.

Testing and Refining the Dataset

We ran all of our testing with Weka, using several of the built-in algorithms we learned in class and testing with both 10-fold cross validation and a test set of the most recent 2 years of data (2013 and 2014). We first ran it on decision trees, because the resulting tree gave us a lot of insight regarding the attributes with the highest information gain. However, we decided against exclusively pursuing decision trees, due to the rigidity of the splits (each item either falls in one bin or the other every time). For these reasons, we decided to perform additional testing using multilayer perceptrons, bayes nets, and logistic regression, predicting neural nets to provide the best results. One notable observation at this point was how 10-fold cross validation and the test set provided similar results for accuracies, meaning we weren't running into overfitting issues with our dataset.

Once this round of testing was completed, an additional attribute was added to the dataset, stating whether or not the team made the playoff in the current year. With this addition, as expected, the decision tree algorithm decided to split on the current year playoff attribute first, as it gave the new highest information gain. However, decision trees still gave lower accuracies than other algorithms, so we proceeded with other algorithms due to each of their unique abilities to assign weights to each statistic to provide the capacity for a wider spectrum of results.

Results and Conclusions

After testing our data with several algorithms, leaving 2013 and 2014 out for use as a validation set, we had the best results with the “Logistic” algorithm in Weka. This algorithm, which performs a logistic regression on the data, ended up classifying 46 out of 64 teams correctly for an accuracy of 71.875%. Most other algorithms produced accuracies in the low 60% range, which only made them trivially better than the ZeroR algorithm, which classified every team as not making the playoffs for a flat 62.5% accuracy. Our predicted “best algorithm,” multilayer perceptrons, was even worse and produced an accuracy of 54.6875%. Each of the Weka results for these three algorithms can be found in *Appendix B.* These results lead to the conclusion that, while regular season results are important, offseason factors like coaching turnover, the Draft, and free agency may be more important. Additionally, we initially got higher accuracies with the last 2 years in the training set using 10-fold cross validation, but we decided to leave 2 years out for a test set to combat overfitting, and our final model didn’t show signs of being overfitted to the training set at all.

Future Testing

The best accuracy we got for our dataset was 71.875%, which is decent considering we’re trying to predict the extreme variability of professional football. However, there are several steps that can be taken in future studies to improve on this value. As mentioned in the results section, there are several additions that could be made to our attributes, such as coaching changes, draft picks, and injuries to name a few. To get a clearer picture of who will make the playoffs this coming season, additional features not directly related to statistics would need to be included in the algorithm. With these additions, new potential ways of training and testing could come to light that result in even higher percentages.

Appendix A: Dataset and Attributes

Below is list of our attributes. All rows are compiled over a 16-game regular season, so playoffs are not included.

PFD - total points allowed on defense
YdsD - yards allowed on defense
PlyD - number of defensive plays
Y/PD - yards allowed per play on defense
TOD - turnovers forced
FRD - fumbles recovered
1stDD - 1st downs allowed
CmpD - completions allowed
PassAttD - pass attempts faced on defense
PassYdsD - pass yards allowed
PassTDD - pass touchdowns allowed
IntD - interceptions
NY/AD - net passing yards per opponent's attempt
Pass1stDD - passing first downs allowed
RushAttD - rush attempts faced on defense
RushYdsD - rush yards allowed
RushTDD - rush touchdowns allowed
Y/AD - rush yards/attempt allowed on defense
Rush1stDD - rushing first downs allowed
ScPercentD - percentage of time opponent scored while you were on defense
TOPercentD - percentage of time defensive drives ended with turnovers
PF - points scored on offense
Yds - offensive yards
Ply - offensive plays
Y/P - yards gained/play
TO - offensive turnovers
FL - fumbles lost
1stD - 1st downs
Cmp - passing completions
PassAtt - passing attempts
PassYds - passing yards
PassTD - passing touchdowns
Int - interceptions thrown
NY/A - net passing yards/attempt
Pass1stD - passing 1st downs
RushAtt - rush attempts
RushYds - rushing yards
RushTD - rushing touchdowns

Y/A - yards per rushing attempt
Rush1stD - rushing 1st downs
ScPercent - scoring percentage on offense
TOPercent - turnover percentage on offense
PlayoffsThisYear - did the team make the playoffs the season this data is from?
Playoffs - did the team make the playoffs the following season?

Appendix B: Weka Results

```
=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances          40               62.5    %
Incorrectly Classified Instances        24               37.5    %
Kappa statistic                          0
Mean absolute error                      0.4688
Root mean squared error                  0.4841
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances               64

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         1         0.625     1         0.769      0.5       No
                 0         0         0         0         0          0.5       Yes
Weighted Avg.    0.625     0.625     0.391     0.625     0.481      0.5

=== Confusion Matrix ===

  a  b   <-- classified as
 40  0 |  a = No
 24  0 |  b = Yes
```

Figure 1: ZeroR Weka results, this acts as a baseline for our training by predicting all teams to miss the playoffs

```
=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances          35               54.6875 %
Incorrectly Classified Instances        29               45.3125 %
Kappa statistic                          0.072
Mean absolute error                      0.4527
Root mean squared error                  0.6232
Relative absolute error                 96.5727 %
Root relative squared error            128.7355 %
Total Number of Instances               64

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.575     0.5       0.657     0.575     0.613      0.543     No
                 0.5       0.425     0.414     0.5       0.453      0.543     Yes
Weighted Avg.    0.547     0.472     0.566     0.547     0.553      0.543

=== Confusion Matrix ===

  a  b   <-- classified as
 23 17 |  a = No
 12 12 |  b = Yes
```

Figure 2: Weka Multilayer Perceptron results, initially our prediction for the best algorithm, neural nets actually performed worse than ZeroR

```
=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances          46               71.875  %
Incorrectly Classified Instances        18               28.125  %
Kappa statistic                          0.3571
Mean absolute error                      0.4269
Root mean squared error                  0.4665
Relative absolute error                 91.0492 %
Root relative squared error             96.3651 %
Total Number of Instances               64

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 0.875     0.542      0.729     0.875     0.795      0.634    No
                 0.458     0.125      0.688     0.458     0.55       0.634    Yes
Weighted Avg.    0.719     0.385      0.714     0.719     0.703      0.634

=== Confusion Matrix ===

  a  b   <-- classified as
 35  5 |  a = No
 13 11 |  b = Yes
```

Figure 3: Weka Logistic results: running a logistic regression results in the best performance of the algorithms we tested

Appendix C: Who did what

Dylan collected and organized the dataset, and ran all of our Weka tests

Ryan created and organized the final website, and wrote the extended abstract and the final report with help from Steven and Dylan

And Hamilton wrote... the OTHER FIFTY-ONE