**Styles Make Fights: Past Plays as a Predictor of Future Plays**

Like most team sports, football teams change up their rosters every year. Because of this, it would be easy to think of the 2017 Patriots and 2018 Patriots as two entirely different teams. But is there some core playstyle that carries over through the years? Can 2017 Patriot plays be used to predict 2018 Patriot plays? To answer this question, we looked to our data.

One of the most interesting pieces of our data is the “play\_type” column. This tells whether each play was a pass, run, punt, quarterback spike, quarterback kneel, extra point, kickoff, field goal, or no play. Seeing the wealth of information available in our csv files, we wondered: could we write a program to predict a team’s plays in various conditions based on their plays in previous years? predict\_plays.py is our attempt to do so.

The first step in answering this question was to filter the data. The csv files contain 231 columns, but most of these columns are not relevant to the question. Besides the “posteam” (the team making the play) and “play\_type” columns, only columns containing numbers were examined. This is because many training models require all values to be numeric. The only column that seemed useful but did not fit this criterion was the “defteam” column, which contains the defending team. After combing through the data for information that would be useful, the choice was narrowed down to seven columns: “down”, “game\_seconds\_remaining”, “quarter\_seconds\_remaining”, “yardline\_100”, “ydstogo”, “score\_differential” and “timeout”. The last column, “timeout”, is a binary variable displaying whether or not a timeout was called. After testing various combinations of these seven columns, the “timeout” column was removed from the DataFrame, because it was hurting the results. This is perhaps because this column is discrete where most others are closer to continuous, so changes from 0 to 1 are weighed more heavily in the models than they should be. A similar point could be made about the “down” column, which gives a number from 1 to 4 to tell which down the play occurred in. Intuitively though, it is easy to see that downs play a bigger part in play selection than timeouts do. The results confirm this; when down is not included, the model’s test scores are much lower. Plays with 0 yards to go were filtered out, since the options available there are different than the options available at all other points in the game. Any row with a null value was also filtered out.

The next step was to select a choice of model to train and test the data. Three options were considered: Gaussian NB, K Nearest Neighbours, and Random Forest. After much testing, the Random Forest model with n\_estimators=300, max\_depth=9 and min\_samples\_leaf=3 was found to perform best.

Using 2017 Patriots data for training, the model was able to successfully predict 2018 Patriot plays with an accuracy of 68.77177949157384% averaged over 30 trials. This is a decent accuracy score. But is the model really learning the Patriots’ playstyle, or is it just learning how to play football? Would the accuracy score be any different if we were to test the model against all NFL plays? It turns out the answer is yes. Using 2017 Patriots data for training, the model was successfully able to predict general 2018 NFL plays with an accuracy of 66.35839897244102% averaged over 30 trials.

So past Patriots plays are better predictors of future Patriots plays than they are of general future NFL plays. This gives some evidence that team playstyles carry over year to year. But what would happen if the training data was pulled from all 2017 teams rather than only the Patriots? How well would this data predict 2018 Patriots plays? Using 2017 general NFL data for training, the model was able to predict 2018 Patriots plays with an accuracy of 68.16052556412454% averaged over 30 trials. This is still lower than the score for 2017 Patriots plays predicting their 2018 counterparts, so this also suggests that team playstyles carry over.

For a bit of extra interest, since it only took one extra line of code, predict\_plays.py also calculates how well 2017 general NFL plays can predict 2018 general NFL plays. The answer is “quite well”. The accuracy for these conditions was 70.04221680738937% averaged over 30 trials. It is interesting that general NFL predictions with general NFL data are more accurate than team-specific predictions with team-specific data. This implies that although styles may carry over for teams, these stylistic differences average out when looking at the NFL as a whole.

The numbers for these prediction scores are close enough that some statistical tests should be done to determine whether or not the difference is meaningful. This analysis is done in plays\_anova.py. The results of 30 trials were manually appended to four lists – one for each accuracy score. You can see the commented-out print statements in predict\_plays.py that provided the append statements (yay for pseudo-automation). Performing a normality test on these lists provides some interesting results. Three of these four lists pass the normality test with flying colours, with p-values of around 0.4, which 8 times the cut-off point of 0.05. Because of this, we fail to reject the null hypothesis that these lists are not normal. Therefore, these lists are probably normal. The list containing the scores predicting 2018 Patriots plays using 2017 general NFL data is a different story though. This list has a p-value of 0.023355820245938876. This means that we fail to reject the null hypothesis that this list is not normal. This is a shame, because this is one of the two groups that we were most interested in looking at. Despite this, some analysis can still be done.

Looking at the 30 sets of prediction scores, NE\_NE (Patriots 2017 predicting Patriots 2018) is higher than NFL\_NE (general NFL 2017 predicting patriots 2018) 26 times out of 30. NFL\_NE is higher three times, and the two are tied once. This suggests that the difference in averages between these two scores is not due to chance. Additionally, 30 samples are enough to follow the patented Greg Baker “it’s probably normal enough” school of thought. So we conducted a Tukey Honestly Significant Difference test to see if the difference between each pair of these four averages was meaningful. The Tukey HSD test found that all six pairs of averages had a meaningful difference. We can now have more confidence that playstyle carried over for the Patriots between 2017 and 2018. But does this apply to the Patriots in other years? And does it apply to other teams?

Looking in depth at other years and other teams would have been too time consuming to do for this project. Some quick eye test analysis with a few other cases will have to do. Training the model with Patriots plays from before 2017 actually predicts 2018 Patriots plays worse on average than when the model is trained with general NFL plays from before 2017. This suggests that team playstyle might not actually carry over for very long. The model also gave mixed results when testing years other than 2018 against training from the previous year (for example, testing 2010 plays against 2009 training). In some of these cases, Patriots data provided better results, but in other cases, general NFL data was preferred. The other teams that were tested in the model (“ATL”, “CAR” and “SEA”) each performed better when trained with general NFL data. So the model telling us that the Patriots’ playstyle carried over between 2017 and 2018 actually seems to go against the trend. Perhaps the prediction being better in this scenario is a fluke after all. Or maybe the Patriots really did retain more of their playstyle between these years than is usual for an NFL team. A more dedicated statistician might look at their roster changes over the years to see if their team changed less between those years, or if they had a change of coaches in other years.

Perhaps this question should have been narrowed. Plays like “pass” and “run” are very broad, and there is actually a lot of room for playstyle within those categories. A run from the Seahawks may look very different than a run from the Eagles. These subtleties, however, were not differentiated in our data. Sports analysts attempting to answer this question in the future should find a way to differentiate between different types of plays within the same category. These future analysts are welcome to use our framework to help them answer the question. Until then, however, the question of team playstyle carrying over from year to year in the NFL will remain inconclusive.