Section 1: Importance

In the modern NFL, play-calling is a strategic battle, with "Play Action" pass plays increasing yardage gained by 1.7 yards per attempt in comparison to a standard drop-back in the most recent 2024/2025 NFL season (Griffis, 2025). In previous years when calling "Play Action" pass plays, play success rate improved 5% (Sharp, 2019). However, not all "Play Action" plays are created equal, their success is influenced by defensive alignment, coverage schemes, and pre-snap tendencies. With the rise of pre-snap motion being used 53% in 2024 the goal is to deceive the defense (Griffis, 2025). When defenders fall for "Play Action" fakes, they move out of position for defending the pass and create open areas of the field for the QB to throw. This is most effective in the secondary level, across the middle, and deep parts of the field (Hermsmeyer, 2019). The understanding of this provides an opportunity to analyze and predict when to call a "Play Action" pass, and how effective it can be. What will be shown in the analysis utilizes supervised data mining techniques such as Random Forest Decision Trees to predict when a "Play Action" play will be called within an NFL game. By performing these techniques; Principal Component Analysis (PCA), and K-Nearest Neighbors (KNN) to classify "Play Action" efficiency with game situational variables and defensive coverages. With the results of this analysis NFL coaches, analysts, and scouts can make informed decisions on when and where "Play Action" is most effective. The project doesn't just identify trends, its goal is to create actionable insights for play selection to gain a competitive advantage in professional football.

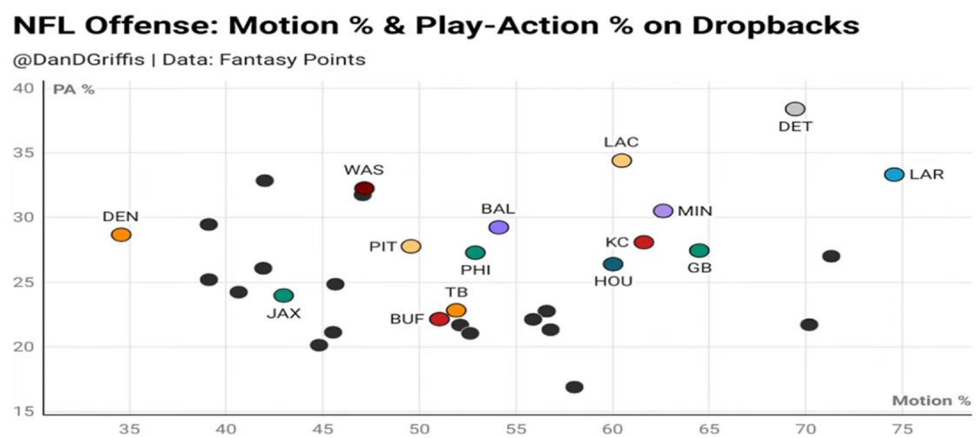


Figure 1. The amount of Motion and "Play Action" pass used in the 2024 NFL season (Griffis, 2025).

Section 2: Statistical Methodology

The definition of "Play Action" in American football is a deceptive strategy where a quarterback fakes a handoff before passing the ball (Anon, 2025). This project utilized data from the NFL's 2025 Big Data Bowl, which consists of weeks 1-9 of the 2022 season, to conduct a two-part analysis focused on "Play Action" passes. The first objective was aimed to develop a supervised learning model capable of predicting whether a play would be a "Play Action" pass based on pre-snap variables. The second objective was to determine whether game situational factors such as down, distance, field position, and defensive coverage could predict the success of "Play Action" pass.

"Play Action" plays accounted for only 17% of the dataset, creating a class imbalance that required specific handling techniques (see Appendix A.2). This involved categorizing down and distance, evaluating defensive coverage schemes, and incorporating pre-snap win probability to assess game-state influence. The yards-to-go variable was tested in both categorical and numerical formats but negatively impacted model accuracy and was ultimately removed. Several models were explored, beginning with a Decision Tree using Gini Index, which yielded high error after the first split and lack of generalization suggesting the model might be too sensitive to small variations in the data, a sign of overfitting. A Random Forest model was then implemented to help mitigate overfitting by training on multiple decision tress and to help stabilize the model. This model improved performance but had low specificity by over-predicting "Play Action" plays, due to high class imbalance. To then mitigate class imbalance, class weighting and the reduction of trees were then tested, yielding minor improvements in specificity (from 0.05 to 0.08), though the model remained moderately effective. Attempts to enhance the predictive power using a Boosting technique to improve model performance by focusing on the errors made in previous trees produced. Adaptive Boosting or Gradient Boosting techniques would have placed more weight on the misclassified "Play Action" plays, gradually improving the models ability to identify them. The technique will be implemented in future iterations because of computational constraints. The technique had taken 32-hours to run before abandoning due to time constraint. Given balance between interpretability and predictive power, Random Forest was selected as the final model. Model performance was assessed through Receiver Operating Characteristics (ROC) curve and Area Under the Curve (AUC), and a cumulative lift chart to show how the model prioritized high-probability situations (see Appendix B).

The second phase of this analysis examined "Play Action" pass success based on pre-snap conditions. Success was initially measured using typical standard measure, Expected Points Added (EPA), but was later refined to a yardage based success metric to better reflect situational effectiveness. A play was considered successful if it gained at least 40% of the required yards on 1st down, 60% on 2nd down, and 100% on 3rd and 4th down (Seth, 2020). This adjustment provided a structured success measure that accounted for game context. To reduce dimensionality, PCA was applied, condensing pre-snap variables into four principal components representing game momentum (win probability shift), scoring differential, down and distance pressure, and field position influence. Defensive coverage types and man/zone schemes were converted to categorical variables and represented their relative function to be compatible with classification models (see Appendix C).

Given the binary nature of the prediction task, KNN was selected because of specific game situations and defensive coverages that had led to a successful "Play Action" play in the past. The model was trained using 5-fold cross-validation and tuned across different k-values to optimize performance. To prevent scale-related biases in the distance calculations, continuous variables were standardized before training. Model performance was assessed through ROC curve and AUC, alongside decile-wise lift charts to evaluate the models ability to rank successful plays (see Appendix D).

Section 3: Results and Interpretation

Prediction "Play Action" Passes

The final random Forest model achieved an 83% accuracy rate, demonstrating high sensitivity but low specificity. While the model effectively identifies "Play Action" plays, it frequently over-predicts them, leading to an AUC of 0.5934 only marginally better than random guessing (Appendix B). This results suggests that the model struggled to distinguish "Play Action" from standard passes and runs. Analysis of Feature importance revealed that defensive coverage, run-pass options, and pre-snap win probability were the strongest predictors, while yards to go oddly had a negative impact on accuracy. This indicates that "Play Action" decisions are driven more by defensive alignment and game flow rather than traditional down and distance factors (see Appendix A).

Further evaluating the cumulative gains chart showed that the model ranked "Play Action" plays only slightly better than random selection. The ROC curve further confirmed weak classification power, with results suggesting that "Play Action" is a strategic coaching decision rather than one dictated purely by pre-snap variables (Figure 2). The lack of detailed formation and motion data possibly limited the models ability to recognize key tendencies that influence play selection.

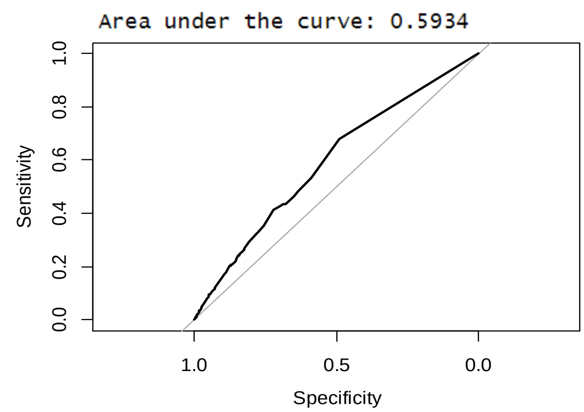


Figure 2. The Area under the curve and ROC curve of the Random Forest model performance.

To improve predictability, future work should incorporate offensive formation details, such as personnel grouping, pre-snap shifts, and motion tendencies. These elements could provide critical context that reflects real-world coaching and decision-making. While this analysis establishes a baseline understanding of how defensive coverage and in game factors influence play calling, further adjustments could give stronger insights to "Play Action" predictability (see Appendix B).

Predicting "Play Action" Success

The best-performing KNN model for predicting play success achieved a validation accuracy of 55.5%, slightly better than random guessing but still relatively low for a predictive model. The ROC curve produced and AUC of 0.52, showing minimal predictive power in differentiating between successful and unsuccessful "Play Action" plays. The curve closely followed the reference line, reinforcing the models difficulty in predicting successful plays (see Appendix D).

To reduce dimensionality, PCA was used on game situation variables. The first four principal components (PC1-PC4) were retained, explaining 98.18% of the total variance in the dataset. These components captured the following game context: PC1 (Game Momentum): strongly influenced by pre-snap win probability, distinguishing between plays where the home or visiting team had an advantage. PC2 (Scoring Differential): captured differences in team scores, reflecting how the current score might impact play-calling. PC3 (Down and Distance Pressure): represented the relationship between yards to go and absolute field position, showing how much pressure an offense was under. PC4 (Field Position Influence): highlighted how a team's location on the field affected decision making (see Appendix C).

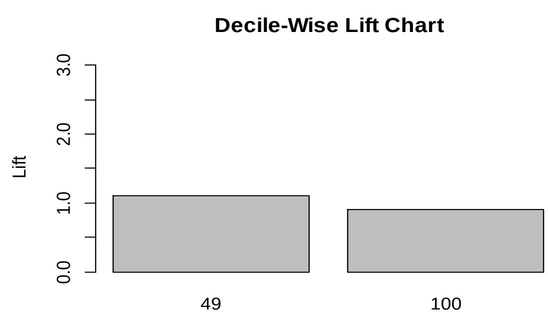


Figure 3. Decile-wise lift chart of the KNN model performance

Further assessment using decile-wise lift chart (Figure 3.) revealed additional model limitations in ranking successful plays. The top half of predictions showed a lift of 1.1, while the bottom half dropped below 1.0, indicating that the model failed to prioritize successful plays above random selection (Appendix D). These findings suggest that pre-snap variables alone do not provide enough predictive strength to reliably forecast "Play Action" success. While game situation and defensive coverage may offer some insight, they do not appear to be strong enough on their own to consistently predict success.

The results highlight that factors such as play design, block execution, and personnel matchups are likely to have a greater influence on "Play Action" success than pre-snap alignment alone. While the model showed some improvement over random selection, it did not reach a level of predictive accuracy to apply practical decision-making. Future work should explore additional features, such as pre-snap motion, which is becoming an increasingly dominant factor in offensive play calling strategies (Griffis, 2025).

Section 4: Alternative Approach

The Random Forest model was ultimately chosen for its balance between interpretability, and handling of categorical features without requiring excessive data transformation. Additionally feature importance analysis provided insights into key drivers of "Play Action". While the models low specificity remains a limitation, the decision tree-based approach allowed for structured evaluation and insights into future improvements. PCA allowed condensing of multiple game situational variables into smaller sets of components, reducing noise while retaining critical game context. KNN was chosen for its ability to naturally group similar plays without requiring strong parameter assumptions about the NFL data. The structured yardage-based success metric improved over traditional EPA measures by capturing the intent behind each play. Although model performance was limited, this approach highlighted the challenges of predicting "Play Action" success based solely on pre-snap factors.

Section 5: Conclusions

The analysis of "Play Action" passes aimed to address the challenge of understanding when "Play Action" is most likely called and what factors contribute to its success. Given that "Play Action" passes have been shown to improve yards per attempt and overall play success rates, optimizing their use can provide an advantage for NFL teams. The first phase of this study demonstrated that pre-snap variables alone are not highly predictive of "Play Action" plays, as the Random Forest model achieved 83% accuracy but suffered from low specificity and an AUC of 0.5934, only slightly better than random guessing. The feature importance analysis revealed that defensive coverage, run-pass options, and pre-snap win probabilities were the strongest predictors, highlighting the state of the game has a larger impact than down and distance.

The second phase examined "Play Action" success, with the best KNN model achieving only 55.5% accuracy and an AUC of 0.52, suggesting that pre-snap factors alone are insufficient for reliable predictions. Decile-wise lift chart analysis further confirmed the models limited ability to rank successful plays above random selection. This indicates that factors such as play design, block execution, and personnel matchups likely play a greater role in "Play Action" effectiveness than pre-snap alignment alone.

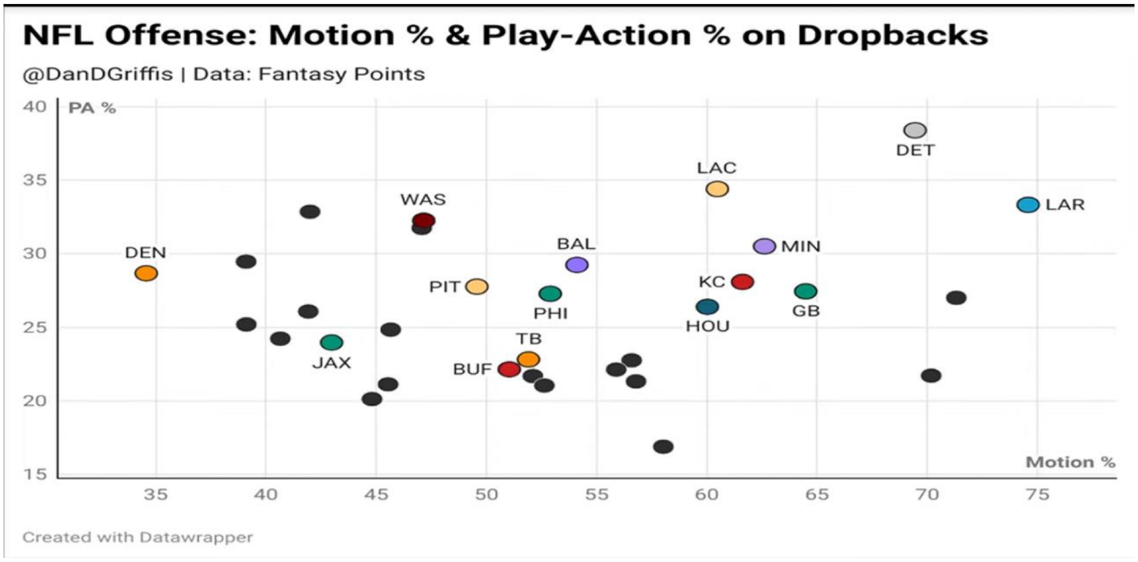
From a business analytics perspective, these results could suggest that if given more detailed NFL pre-snap and formation data, predictive modeling could improve for both "Play Action" occurrence and success. While the models in this study established a baseline understanding of "Play Action" tendencies, future improvements could leverage scouting data, tracking technology, and film analysis to gain deeper insights into real-time play-calling strategies. By refining the data inputs and expanding beyond pre-snap variables, teams can make more informed decisions on when and how to maximize the effectiveness of "Play Action" passing to gain a competitive edge on the field.

Work Cited:

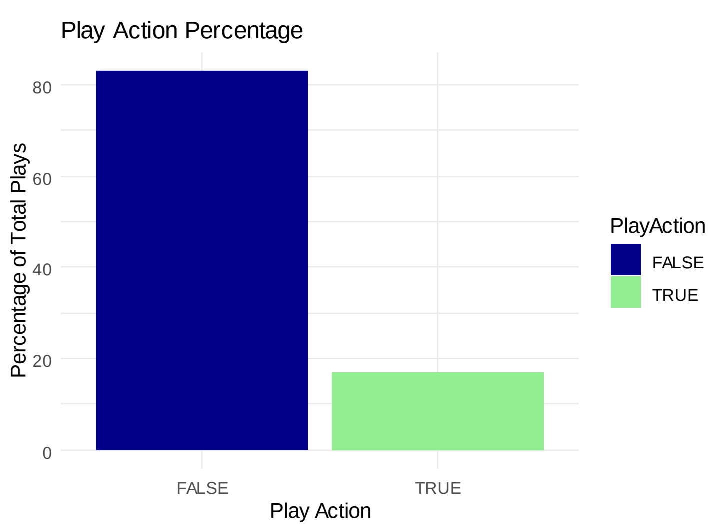
- https://sports.yahoo.com/article/motion-play-action-revolutionized-nfl-234354517.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAKFa3KFdCxEj9NQylrrwsi-BwMMFLvJy50KDRkN_vwwjBluA6kyak_jc8VEkomiINlAd1ZVYIMV2LpR6HDjZseuoZT5FLSh81ZbezgwmzZnd49twlN_jK7v7oCo09EjSqjKm7fgeu1o7sS57cOYJxrzMcOJGkkabvQpp4WO1Y96m (Griffis, 2025)
- "NFL Big Data Bowl 2025." *@Kaggle*, 2024, www.kaggle.com/competitions/nfl-big-data-bowl-2025.
- Seth. "The NFL's Run Gap Secret." *M-FANS*, 15 July 2020, mfootballanalytics.com/2020/07/15/the-nfls-run-gap-secret/. Accessed 7 Mar. 2025.
    - url: https://mfootballanalytics.com/2020/07/15/the-nfls-run-gap-secret/
- Hermsmeyer, Josh. "Can NFL Coaches Overuse Play-Action? They Haven't Yet." *FiveThirtyEight*, 4 Jan. 2019, fivethirtyeight.com/features/can-nfl-coaches-overuse-play-action-they-havent-yet/. Accessed 7 Mar. 2025.
    - url: https://fivethirtyeight.com/features/can-nfl-coaches-overuse-play-action-they-havent-yet/
- "Merriam-Webster Dictionary." *Merriam-Webster.com*, 4 Mar. 2025, www.merriam-webster.com/dictionary/play-action%20pass.

Appendix A

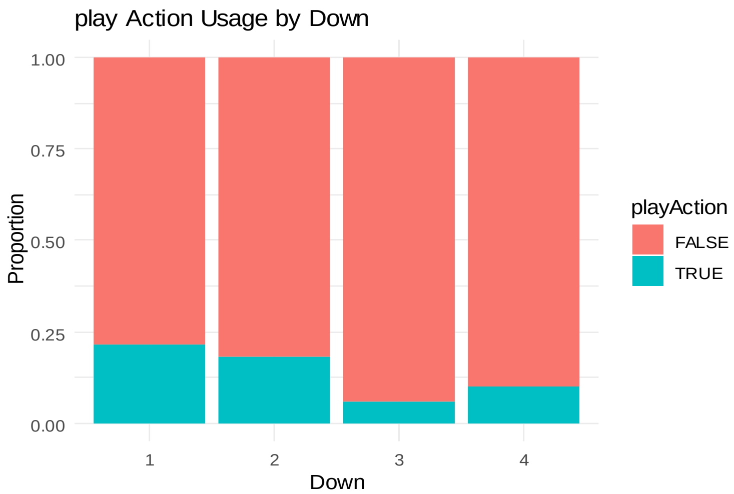
Exploratory Analysis

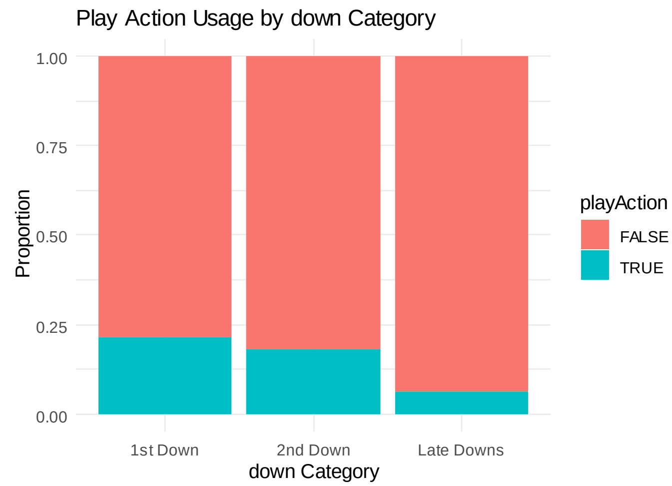


Appendix A.1 This shows the amount of "Play Action" plays and motion that was used in offensive systems during 2024 before the playoffs. (Griffis, 2025)

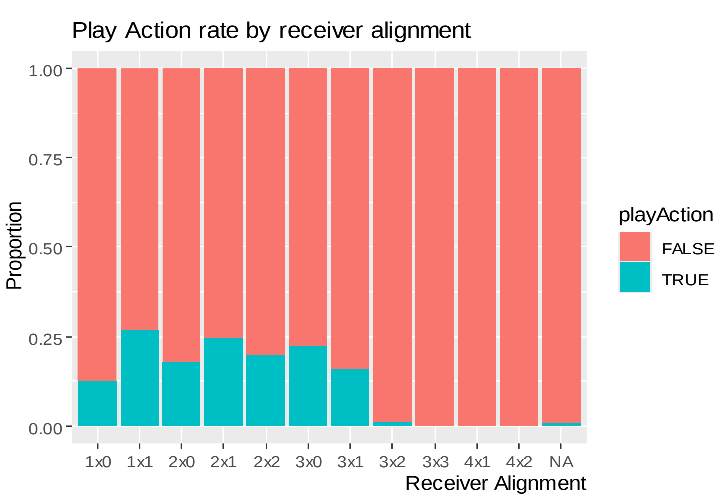


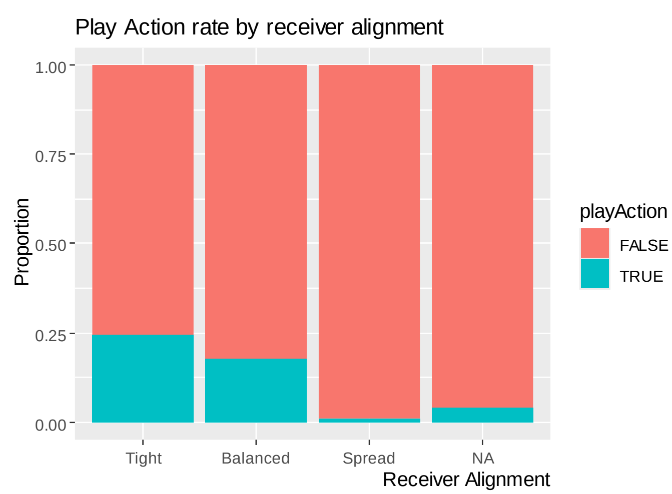
Appendix A.2 The total amount of plays called and the proportion that were plays action in the 2022 season weeks 1-9.



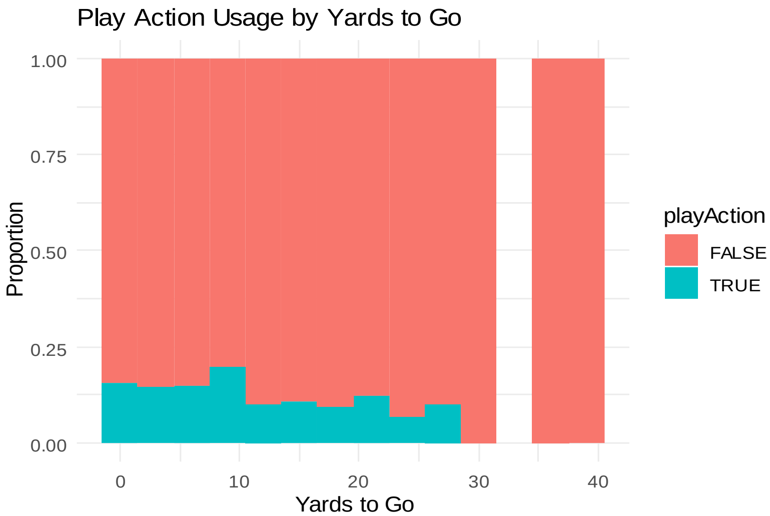


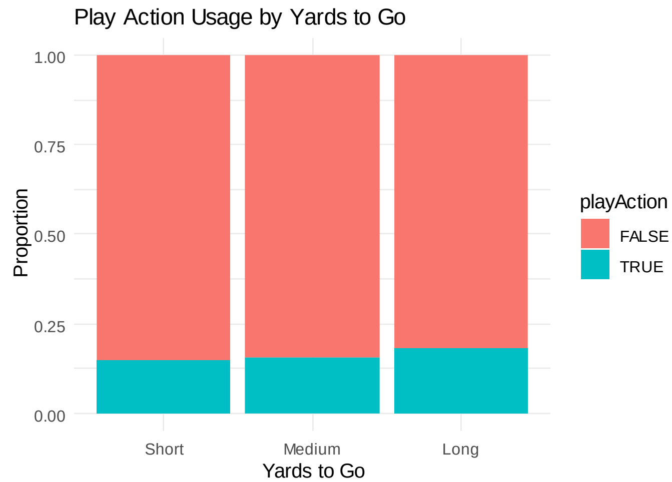
Appendix A.3 The proportion of plays that were "Play Action" called on each down. Since "Play Action" was used less on later downs I combined 3rd and 4th down as one category



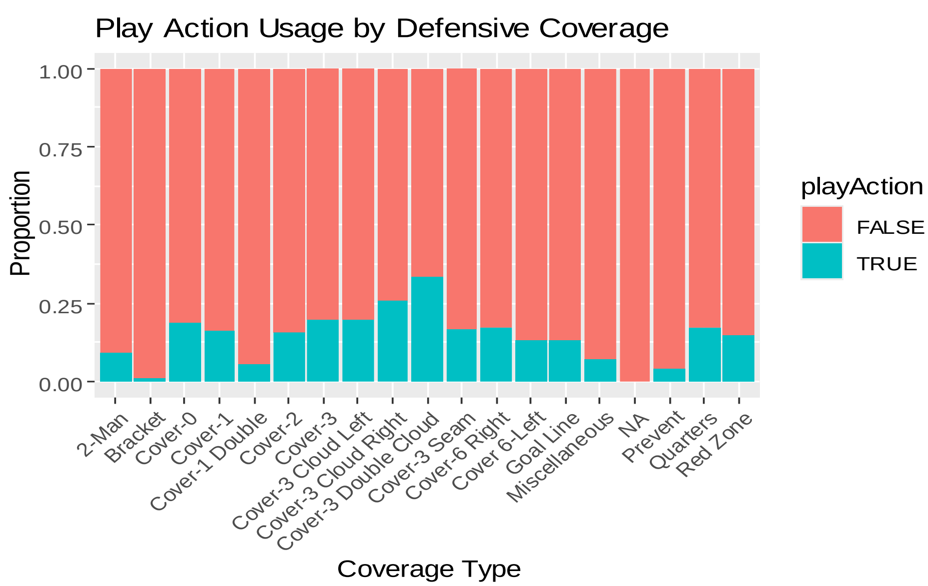


Appendix A.4 The offensive receiver alignment when "Play Action" was called. With the specificity of receiver alignment per play had many different alignment types I grouped them in terms of their position relative to the offensive formation. Tight meaning majority of the offensive personnel is grouped near the offensive line (interior). Balanced there are multiple wide receivers split out. Spread majority of the offensive personnel are not in the interior of the play. NA are miscellaneous alignments not fitting to common structure.

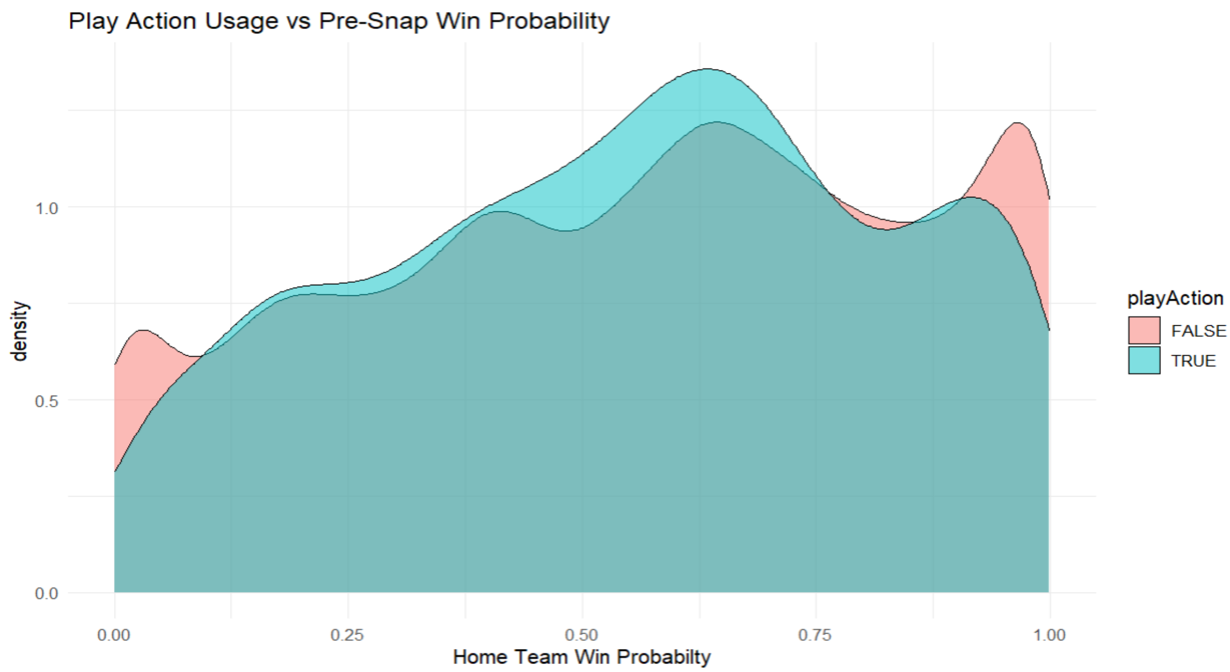




Appendix A.5 When "Play Action" was called in terms of yards to first down or score. Since there were many different situations where "Play Action" was called in yards to go. I categorized the yards into short (< 3yards) Medium (< 9 yards), and Long (> 10 yards).



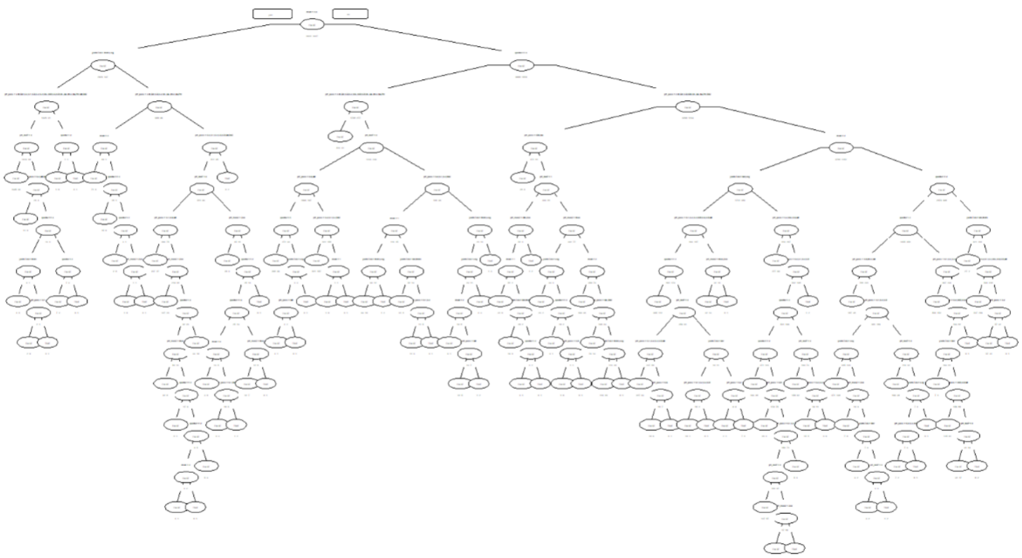
Appendix A.6 The type of defensive coverage that was on the field when "Play Action" was called.



Appendix A.7 The win percentage of the team when they utilize “Play Action”. Exploring if teams in a winning position during the game call “Play Action” more.

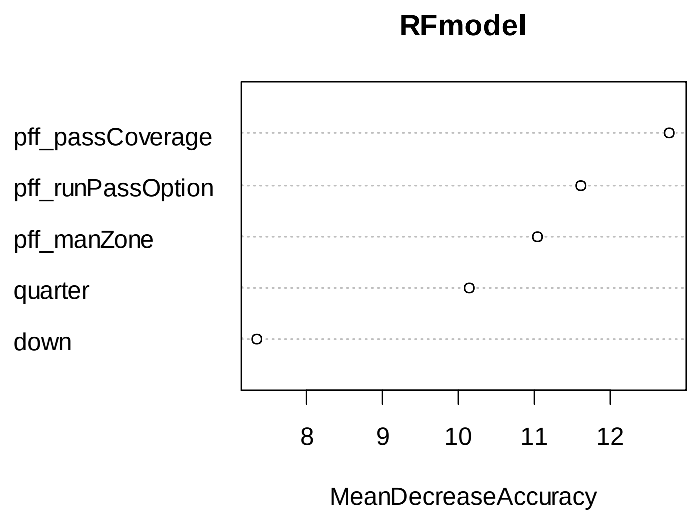
Appendix B

Decision Trees



	CP	nsplit	rel error	xerror	xstd
1	3.0358e-04	0	1.00000	1.0000	0.022446
2	2.4287e-04	6	0.99818	1.0134	0.022564
3	2.0239e-04	21	0.99393	1.0261	0.022676
4	1.9319e-04	37	0.99029	1.0267	0.022681
5	1.7348e-04	75	0.98118	1.0279	0.022692
6	1.5179e-04	82	0.97996	1.0340	0.022744
7	1.2143e-04	90	0.97875	1.0346	0.022750
8	8.6738e-05	95	0.97814	1.0407	0.022802
9	0.0000e+00	102	0.97753	1.0407	0.022802

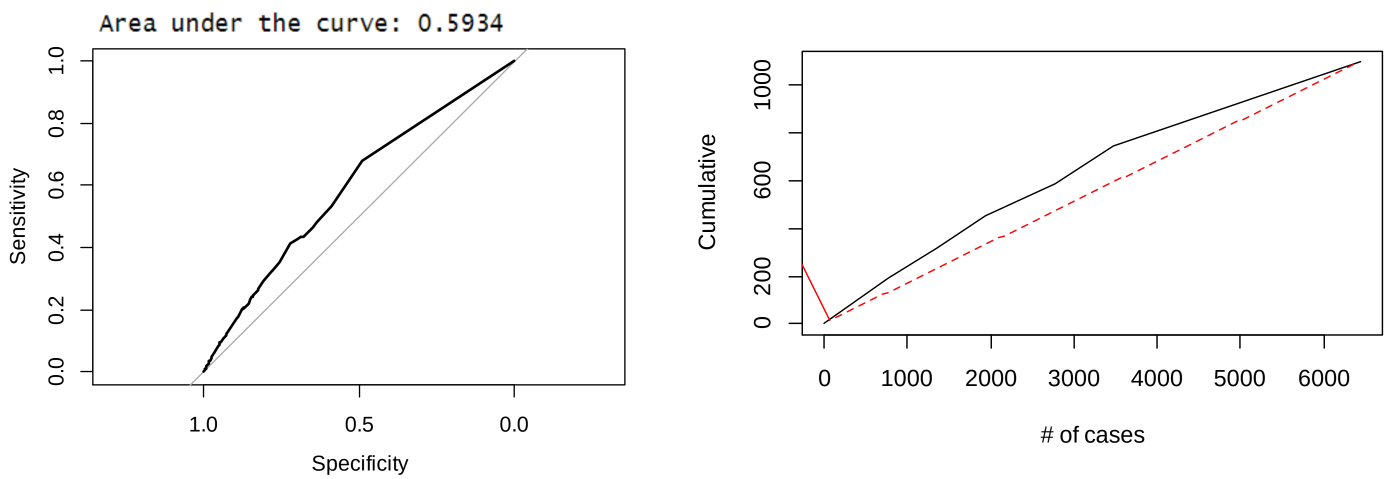
Appendix B.1 This is the first full decision tree. Data was partitioned and split 60/40 for the model to train on. This included Gini index when creating the model. On the right hand side shows the xerror where it never truly went below 1. Thinking this error is stemming from class imbalance leading to overfitting. Ensemble tree was then considered.



Appendix B.2 Executing a Random Forest model I combated the imbalance of data by applying a 0.5 weight to “Play Action” plays to balance out the possible compensation that the model is applying. I also changed the number of trees to train on from 500 to 100 to increase specificity because specificity was low. After tweaking model performance, yards to go negatively impacted model my mean decrease accuracy of -10 so I removed yards to go variable from the model since it seemed to show no relationship, which ultimately make sense when looking at the exploratory model of yards to go and its even spread. I then noticed home and visitor team win percentage per play because it was overpowering decision making of the model, both with a mean decrease accuracy of 30. Indicating that there is possibly a high correlation, as teams are winning the amount of “Play Action” usage increases.

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE TRUE
     FALSE  5234 1053
     TRUE    117   45

               Accuracy : 0.8186
                 95% CI : (0.809, 0.8279)
    No Information Rate : 0.8297
    P-Value [Acc > NIR] : 0.9915

                  Kappa : 0.0289

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.97813
            Specificity : 0.04098
         Pos Pred Value : 0.83251
         Neg Pred Value : 0.27778
             Prevalence : 0.82974
         Detection Rate : 0.81160
   Detection Prevalence : 0.97488
      Balanced Accuracy : 0.50956
```

Appendix B.3 The results of the confusion matrix of the final chosen model after training shows that even with multiple tweaks to increase model performance accuracy was still relatively high but specificity was still extremely low. Indicating that the model had a hard time distinguishing between a “Play Action” play and a non-“Play Action” play.



Appendix B.4 The ROC curve (left-hand side) shows the black curve representing model performance and the grey diagonal line represents a random guess (AUC = 0.5). The models AUC = 0.5934 which means the model performs slightly better than random chance at separating “Play Action” vs non-“Play Action” plays. The cumulative gains chart (right-hand side) shows model performance with the black line and the red-dashed line represent random baseline. This graph shows that the model does rank “Play Action” plays slightly better than random but not by much, overall it has some predictive power, but it isn’t strongly differentiating between “Play Action” and non-“Play Action” plays.

Appendix C

PCA

	PC1	PC2	PC3	PC4	PC5	PC6
yardsToGo	0.008310737	0.03666915	-0.713829392	-0.69930852	-0.001222980	-1.348488e-17
absoluteYardlineNumber	0.017011866	0.04279355	0.700283738	-0.71237701	-0.000993126	1.093331e-18
preSnapHomeTeamWinProbability	-0.651535670	-0.01521773	0.004007091	-0.01215197	-0.274054674	7.071068e-01
preSnapVisitorTeamWinProbability	0.651535670	0.01521773	-0.004007091	0.01215197	0.274054674	7.071068e-01
preSnapHomeScore	-0.254669358	-0.71689758	0.004141691	-0.04597782	0.647356369	-9.044786e-17
preSnapVisitorScore	0.292895183	-0.69456374	0.001009243	-0.03282197	-0.656289052	4.733728e-17

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	1.5281	1.2482	1.0069	0.9919	0.33038	9.368e-16
Proportion of Variance	0.3892	0.2597	0.1690	0.1640	0.01819	0.000e+00
Cumulative Proportion	0.3892	0.6489	0.8178	0.9818	1.00000	1.000e+00

Appendix C.1 This shows the variables that went into the PCA and overall what their weight is within each component. PC1 will represent Pre-snap Win probability (Game Momentum) with the main contributors to win probability. A higher PC1 value means the visitor team has a higher win probability and a lower PC1 value would mean the home team has a higher win probability. PC2 will represent Scoring Differential, main contributors are home score and visitor score. If PC2 value is high then there is higher scoring games and a lower PC2 value would mean lower scoring games. PC3 main contributors were yards to go and absolute yard line number. This will represent down and distance pressure, so the higher PC3 value the closer to the first down or deep in opponent territory the team is, and a lower PC3 value would mean longer distance to convert or the team is backed up near their own end zone. PC4 will represent Opponents field position impact. The main contributors are yards to go and absolute yard line number both with high negative values. This has a different meaning than PC3. This interprets how much the opponent's position affects play-calling. Higher PC4 value would mean a shorter distance to a first down but on opponents side of the field and a lower PC4 would be a longer conversion on own side of the field. Based off of the summary and eigenvalues (table of importance below) PC1-4 will be used to reduce dimensions and still retain 98% of the variance/information from the game situation variables

Appendix D

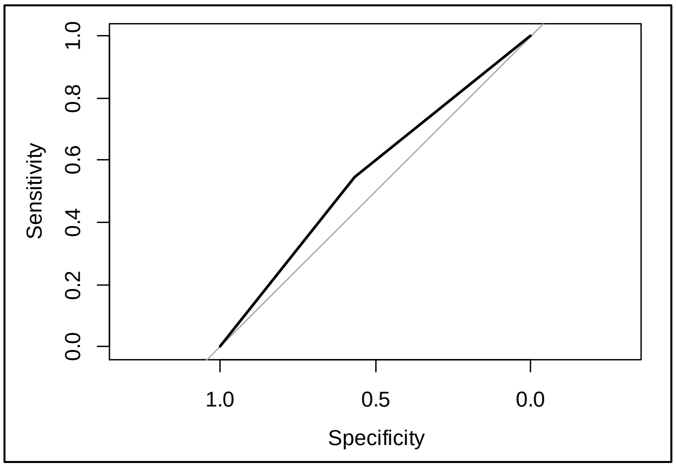
KNN

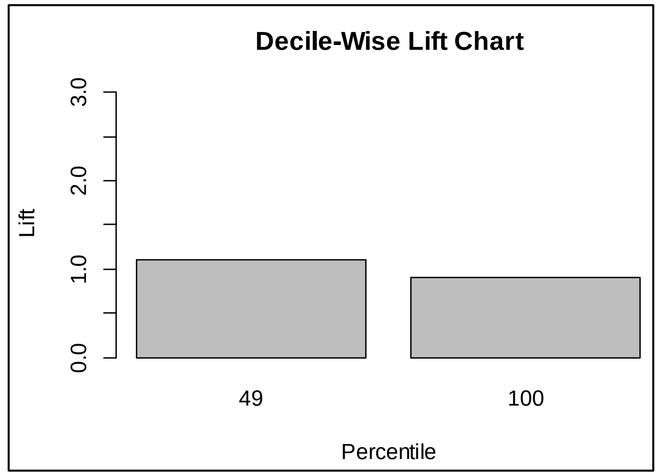
```
  k   Accuracy   Kappa
   1  0.5311804  0.0599239210
   2  0.5161524  0.0288209589
   3  0.5206958  0.0381239159
   4  0.5184158  0.0335401542
   5  0.5161348  0.0273823275
   6  0.5275285  0.0507522352
   7  0.5125150  0.0190822485
   8  0.5184282  0.0315499874
   9  0.5138921  0.0205082521
  10  0.5138880  0.0208040502
  11  0.5129934  0.0172326807
  12  0.5057134  0.0034878456
  13  0.5047908  0.0008750463
  14  0.5111617  0.0126483133
  15  0.5143435  0.0177887377

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was k = 1.
```

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 147 131
         1 113 157

               Accuracy : 0.5547
                 95% CI : (0.512, 0.5969)
    No Information Rate : 0.5255
    P-Value [Acc > NIR] : 0.0923

                  Kappa : 0.1102

 Mcnemar's Test P-Value : 0.2765

            Sensitivity : 0.5654
            Specificity : 0.5451
         Pos Pred Value : 0.5288
         Neg Pred Value : 0.5815
             Prevalence : 0.4745
         Detection Rate : 0.2682
   Detection Prevalence : 0.5073
      Balanced Accuracy : 0.5553

       'Positive' Class : 0
```

Appendix D.1 The data for the KNN model was partitioned to 80:20 training and test split. Results with the final model she showed the highest training accuracy. Cross-validation was used to help the KNN model generalize to new data and help prevent overfitting. In the original model estimated points added (EPA) was used as the success metric (very common with football data) which resulted in a model with 50% accuracy. Changing to a different success metric where 1st downs yards gain was > 40% yards to go, 2nd down yards gained was >60% yards to go, and 3rd and 4th down equaled a conversion resulted in a more accurate model 55% (results above). KNN model overall performs fairly week with an accuracy above random but may not be the best model for predicting “Play Action” success. With a k= 1 suggest that there is overfitting and the model is memorizing the training data rather than generalizing well.

Area under the curve: 0.5553





Appendix D.2 A ROC Curve was selected to use with this model because of the binary classifications and probability scores. The KNN model struggled to differentiate between successful and unsuccessful "Play Action" plays. Slightly better than random guessing for about half the data in the model, then there is diminishing returns after about 50 percent of the data used in model. A decile-wise lift chart was chosen to help distinguish effectiveness of the model in its decision making. This model shows that the top half of predictions is slightly better than random guessing. Overall the model doesn't show real value in predictive power of "Play Action" success.