CREATING A WIN PROBABILITY MODEL: NATIONAL FOOTBALL LEAGUE

Michael Mistarz

Northwestern University, MSDS456: Sports Performance Analytics

GitHub: https://github.com/mistmr7/MSDS456_NFL-WinPobabilityModel-R

July 24, 2025

1. Win Probability Model

In creating a win probability model, I utilized nflfastR to obtain play-by-play data for ten NFL seasons ranging from 2015 through 2024. The data was loaded into a DataFrame, and a column was added for whether the home team won the game and whether the home team currently had possession of the ball. The DataFrame was split into train and test pools through random sampling, with 75 percent of the plays being placed in the training set and 25 percent placed in the testing set. A linear regression and random forest model were created by training on the training set, where wins were being predicted by a combination of down, distance to first down, yard line, seconds left in the game, and score differential. These two models were tested using the testing dataset, and a summary of the important factors for each model was created and an accuracy score for the training model was calculated. This accuracy score was derived by checking if the model predicted each play was run by a winning or losing team and finding the ratio of the amount that it predicted correctly.

The random forest model took several minutes to train, so each step previously listed was repeated for a single season to try to discern what changes improved model accuracy. The final model was rebuilt on the ten-season dataset using the best model from these single season iterations. After trying multiple iterations, including adding estimated points added per play (epa), possessing team, and allowing term interactions including game seconds remaining x score differential, the final model was chosen as the model with the highest accuracy values. This corresponded to a random forest model where wins were determined by down, yards to first down, possessing team, yards to endzone, and score differential. This model was trained on the dataset containing ten seasons worth of data and win probability functions were created to plot NFL games using the model.

2. Single Game Example

After presenting the model to management, they had questions and concerns that the model would not accurately capture the win probability for the team. They asked me to evaluate an NFL game to determine whether my model followed the general swings that were seen in the games. Using the random forest model above, I created a function to assign win probability predictions for the home team based on each play in the game. Each play in the game represents a datapoint in the plot, and the chart can be seen below in Figure 1. This plot represents the win probability for the Chicago Bears Week 1 game against the Tennessee Titans in 2024. I will use the chart to walk through 5 plays or key sets of downs in the game that caused major changes to the model's behavior. When a single play is highlighted, I will go through the description of the play and explain the model's behavior. When a section is highlighted, I will walk through the game situation and explain how the model reacted to it.

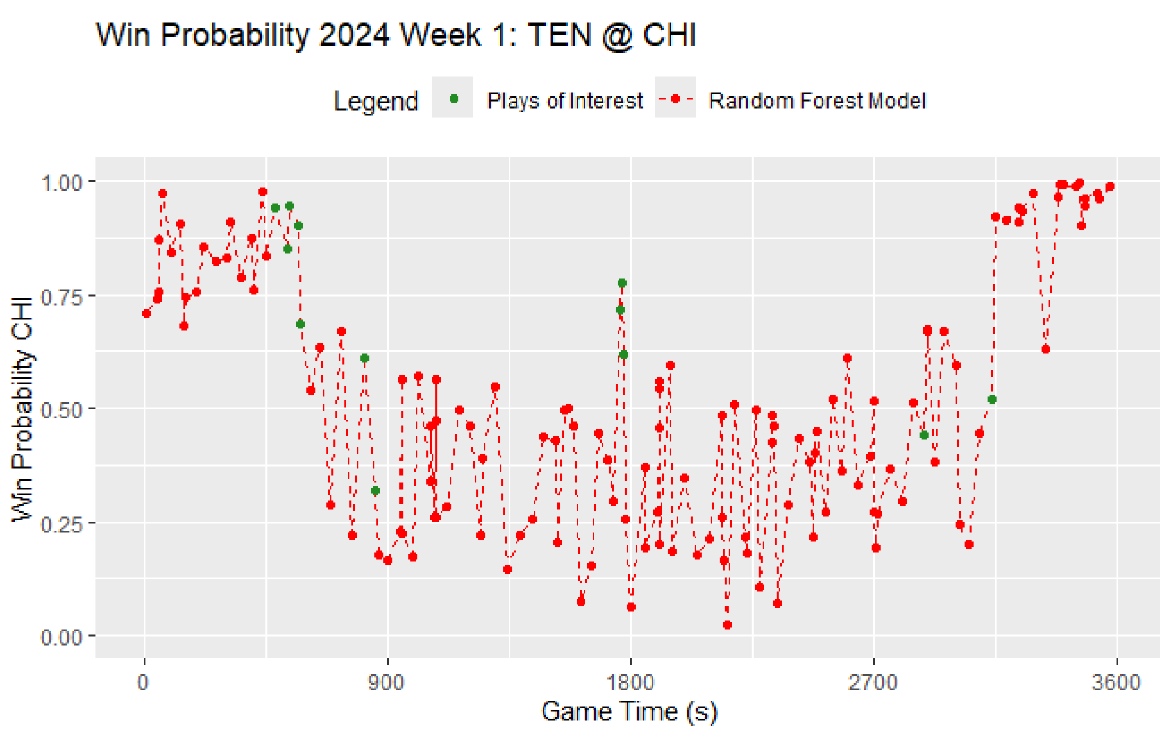

Figure 1: Random Forest win probability model applied to week 1, 2024

First Example:

On Chicago's second drive of the game, seen in the first cluster of green dots, The Bears start with a first and ten situation on Tennessee's 25-yard line. Caleb Williams was sacked for a loss of 19, a screen play was run for another negative 1 yard and then an incomplete pass was thrown. This moved them out of field goal range and forced them to punt, leading to a drop of win probability of 23 percent over those plays.

Second Example:

On Tennessee's last drive of the first quarter, seen in the second pair of green dots, Tony Pollard runs for a first down and then a 19-yard touchdown in consecutive plays to make the score 7-0, leading to a 42 percent drop in win probability by the model.

Third Example:

As Chicago was closing out the half, seen in the third set of green dots right before halftime, The Bears had a first and goal situation at Tennessee's 10-yard line. After a short completion of four yards to set up second and goal from the 6, Chicago had two incomplete passes and was forced to kick a field goal for a 17-3 halftime deficit. Getting held to a field goal in this situation dropped the win probability by 73 percent.

Fourth Example:

With Tennessee up by four points with just under 12 minutes left, as can be seen in the second to last green dot from the right side of the plot, Will Levis dropped back to pass at his own 48 and was sacked. He fumbled the ball, and it was recovered by the Bears at the Tennessee 31. This led to an increase in the win probability for the Bears of 26 percent.

Fifth Example:

Tennessee had the ball with just under 8 minutes to play, as can be seen in the last green dot on the right side of the plot. It was third down and 6 with the Titans up by 1 and Will Levis dropped back and threw a screen pass. The pass was intercepted by Tyrique Stevenson and taken 43 yards for a touchdown. This along with the subsequent two-point conversion gave the Bears their first lead of the game, leading to a 46 percent increase in win probability.

3. Comparing the Model to NFLreadR

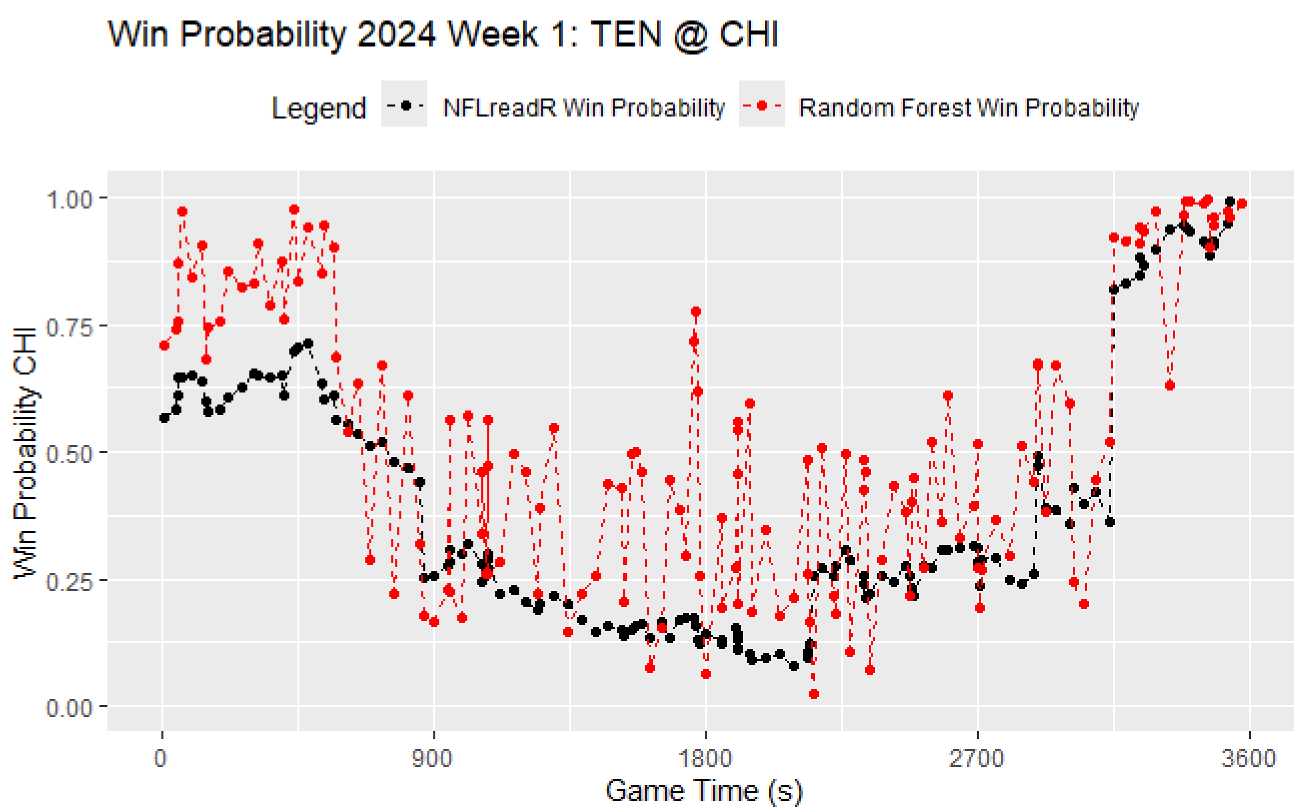

Figure 2: Week 1 2024: Tennessee @ Chicago win probability plots showcasing NFLreadR vs. the random forest model

As can be seen in Figure 2 above, the NFLreadR win probability for the game described in section 2 is plotted alongside the random forest model created for management. There are several key differences that should be highlighted. The first difference, noted in pretty much every plot, is that the random forest model overestimates the probability of the home team winning. According to the play-by-play data over the past ten seasons, the home score was greater than the away score for 54.8 percent of the plays. The average difference in results over those ten seasons was 1.8 points, showing that the home team on average starts the game with a 1.8-point lead over the road team. This is pretty accurately conveyed by the NFLreadR win probability and overestimated by the random forest model, as the random forest model starts closer to 70 percent for the home team.

Another major difference between the two models is the lack of noise in the NFLreadR model compared to the random forest model. The random forest model generally trends in the right direction, but the point-by-point calculations of the win probability show major swings. This does not follow along with the play-by-play data. Some large swings happen when a team punts on fourth down and the punt is caught for a fair catch, and some situations where a team is in long yardage situations do not appear correctly either. Further discussion on these aspects will be covered in the next section.

4. Recommendations, Future Changes, and Limitations

The random forest model created generally trends in the correct direction as the game progresses, but it seems to fail to capture the importance of individual plays to that of the entire game. To try to see where the model currently has issues, I created plots for each of the six variables used to calculate the win probability. I created double plots for each variable, showing how common each variable occurred over the ten seasons (i.e. 1st down, 2nd down, 3rd down, or 4th down) and the corresponding win percentage value applied to the model for each. These plots can be seen in Figure 3 through Figure 8 below.

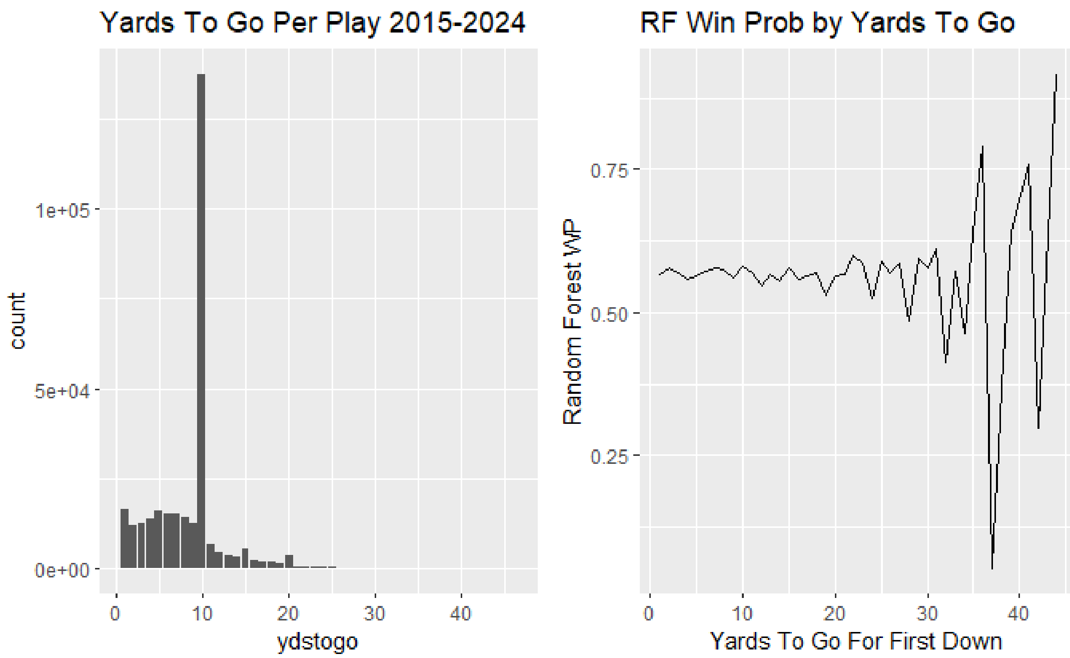

Figure 3: Yards to go until first down count (left) and model calculated winning percentage by yards to go (right).

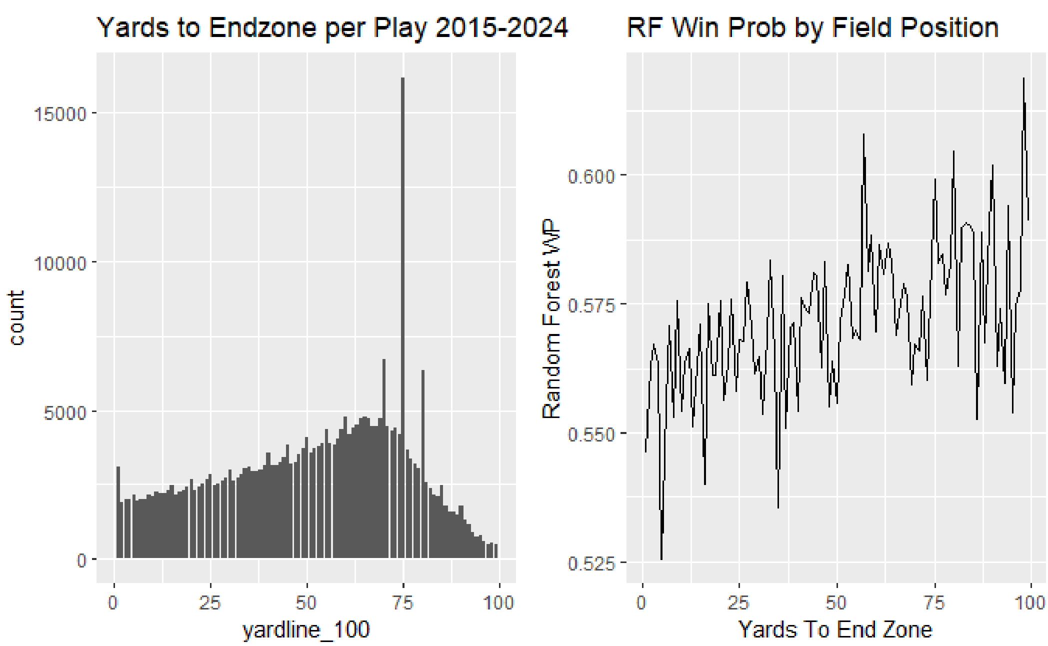

Figure 4: Yards to go until the end zone (left) and model calculated winning percentage by yards to endzone (right).

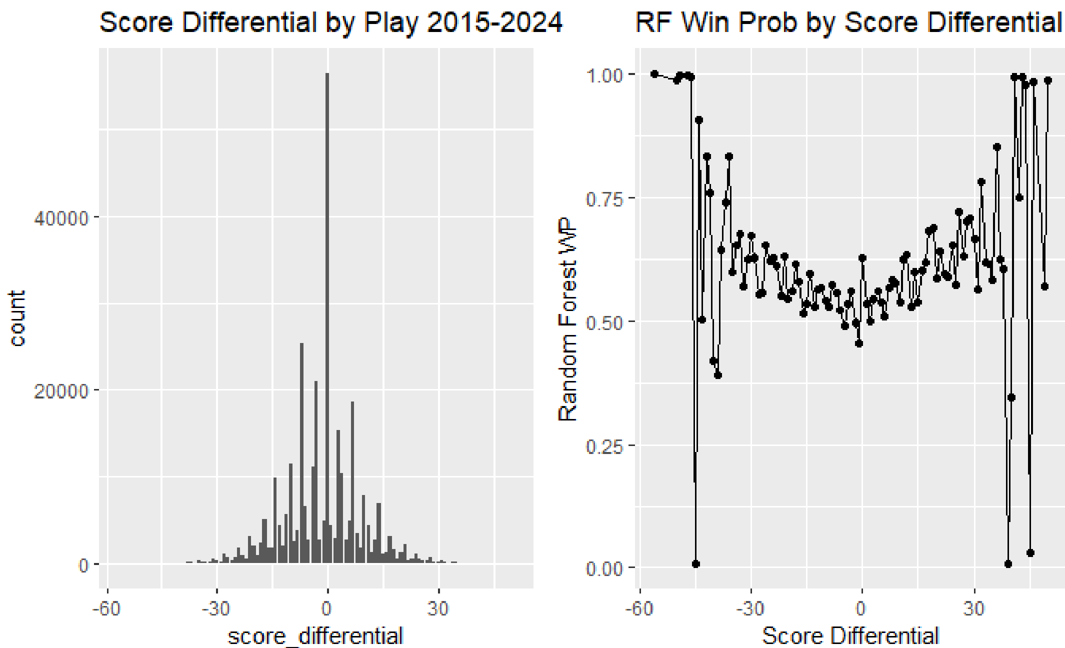

Figure 5: Score differential count (left) and model calculated winning percentage per score differential situation (right).

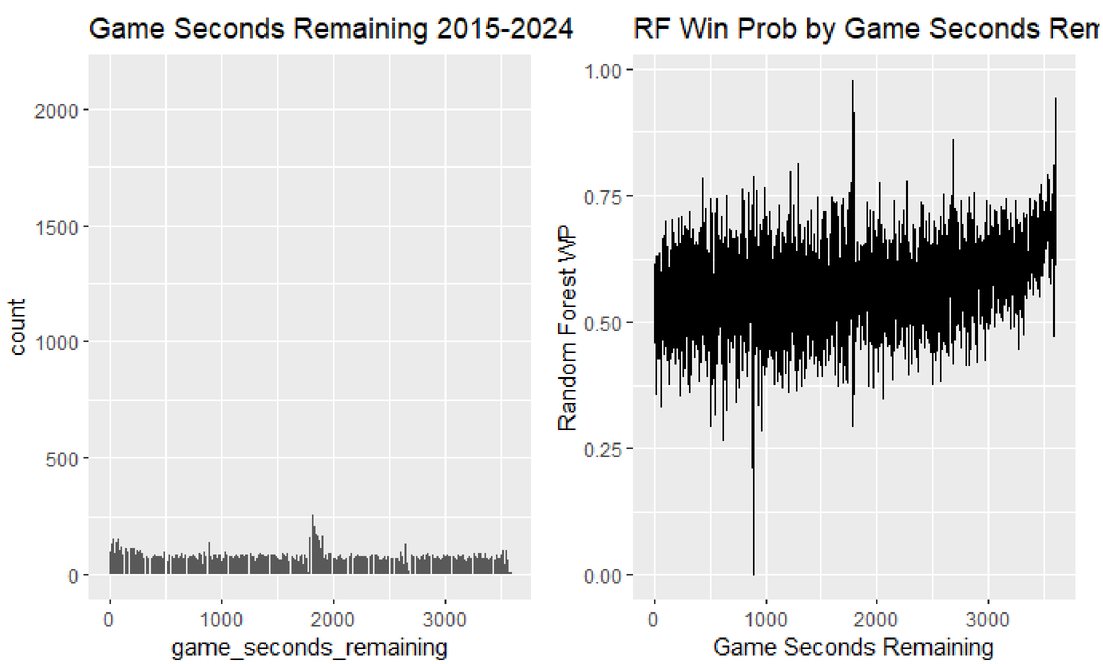

Figure 6: Game seconds remaining play count (left) and model calculated winning percentage by game seconds remaining (right).

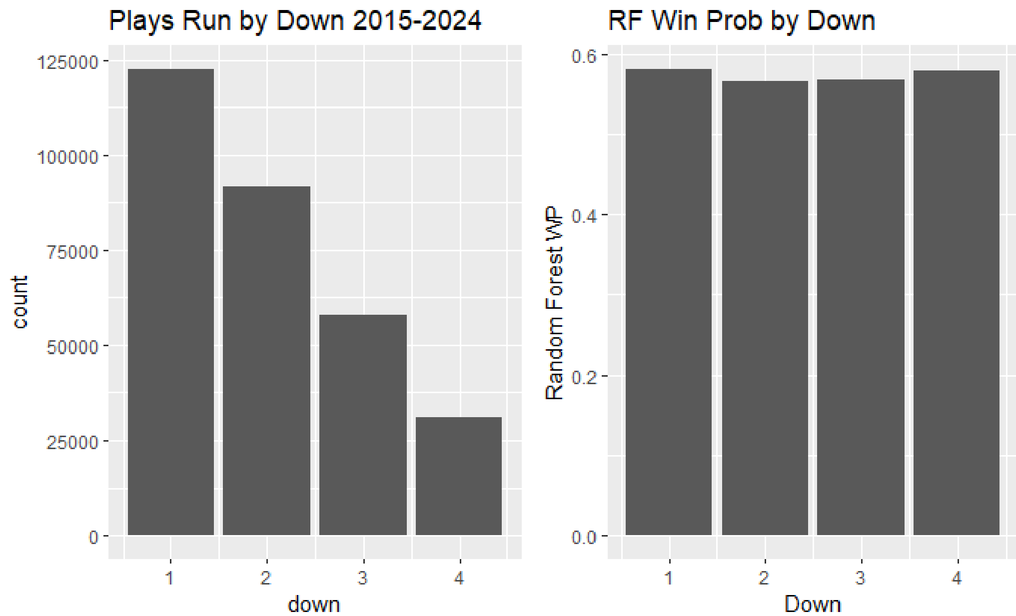

Figure 7: Play count by down (left) and model calculated winning probability by game seconds remaining (right).

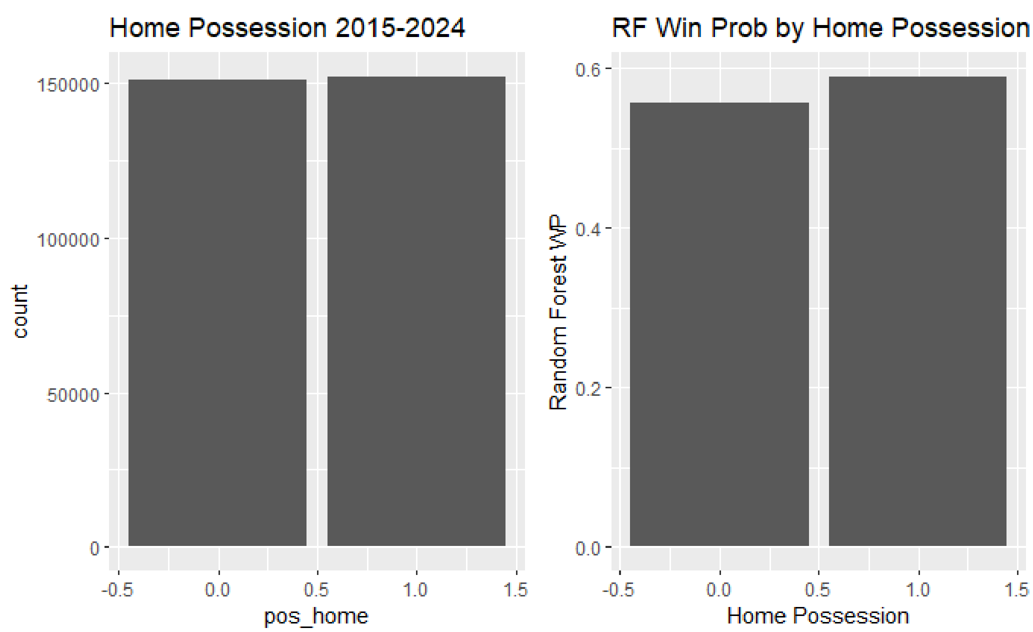

Figure 8: Home possession play count (left) and model calculated winning percentage by home possession (right).

Taking a look at Figure 3, the model clearly struggles when yards to go for a first down are larger than about 20, as the model's calculated win probability gets noisier as yards to go increases. This is likely due to a lack of input values for these regions, and model performance could be improved by treating anything over 20 yards as 20 yards to go. Taking a look at the other yardage option, by field position in Figure 4, the model does not follow intuitive logic, with situations where the possessing team was assigned a higher win probability by being further from the end zone on average. This is clearly a place where the model could use some tweaking. Score differential looks like another place where the model could use some tweaking, though score differential does not assume that it is a score differential for the home team. This could explain the high amount of noise in the figure. The plot was reconfigured with the home team as the reference, and the resulting plot in Figure 9 below appears to show a proper trend. The model likely needs a steeper slope to capture the probability of a twenty plus point comeback, so tweaking could help at the edge cases for score differential.

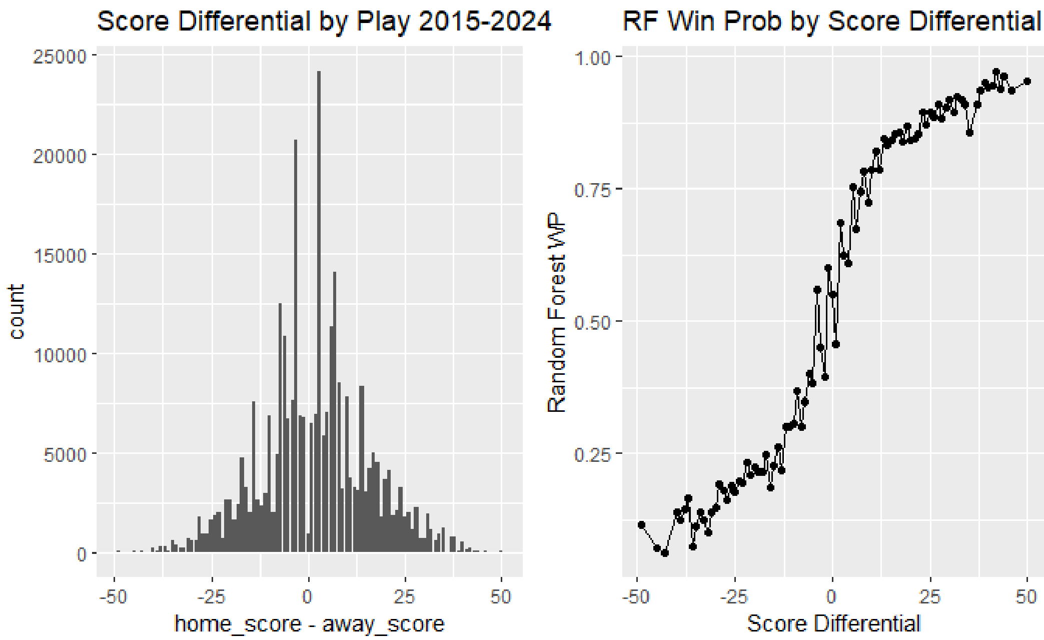

Figure 9: Score differential with respect to the home team (left) and model calculated winning percentage by home score differential (right).

The other three factors appear to be appropriately accounted for. Figure 6 shows how the model applies further value to situations that appear later in the game seems intuitively correct. Figure 7 likely needs tweaking, but that will need further study as well. Intuitively, NFL teams try to get back to first down, so first down should show a higher win probability value than the other downs. Keeping to this logic, 4th down would have the lowest win probability value but for some reason has the second highest value. Attempting to add in yards to go to the down field to join them together could help smooth this out. Figure 8 shows a slight edge in win probability value for the home team, which matches the trend of the home team winning more often.

Because of the noisiness of the model, I do not believe it is an appropriate tool to use to decide specific game situations. I do think this is something the model can account for if tweaked appropriately. Sitting down with the coaching staff and getting an understanding of situations that they want to understand win probability based on their decision making can help build tools to help them in those specific situations. Expected point values can be calculated based on each of the variables above, and tweaking the model to account for historical success rates for different situations could provide the coaching staff with all the data they would need to make those decisions (Baumer, Matthews, and Nguyen 2023). We could also make charts to discuss the probability of making field goals, first downs, or touchdowns from different situations using customized functions (Elmore and Urbaczewski 2025). We could add models to check for field goal success in a certain situation, including factors such as wind direction, temperature, and game situation. There are many ways this project could be improved that could help our teams make better in-game decisions and theoretically lead to more wins for the franchise.

5. Other Analytics Projects

When looking for ways that analytics could be used to help the team, several projects come to mind. Injuries consistently derail seasons for promising teams, and many injuries are unpreventable that occur from in-game situations like chop blocks, falling lineman rolling on an ankle, concussions coming across the field, or even ankle sprains coming down from a contested catch. However, many soft tissue injuries come from overuse and using analytics to help monitor and prevent these would be instrumental to putting the best team on the field. A simple way would be with a continual in-season measure of the countermovement jump (Vald Performance 2024). Figure 10 shows an example of the countermovement jump, and it should be trained daily throughout training camp to develop a baseline for each player. The force plates used in the measurement could be analyzed through computer software to determine fatigue levels for players, allowing the team to tweak practice and training protocols to match the fatigue level of the player. We could collect data from players that sustain injuries and continually improve our injury prevention process over time.

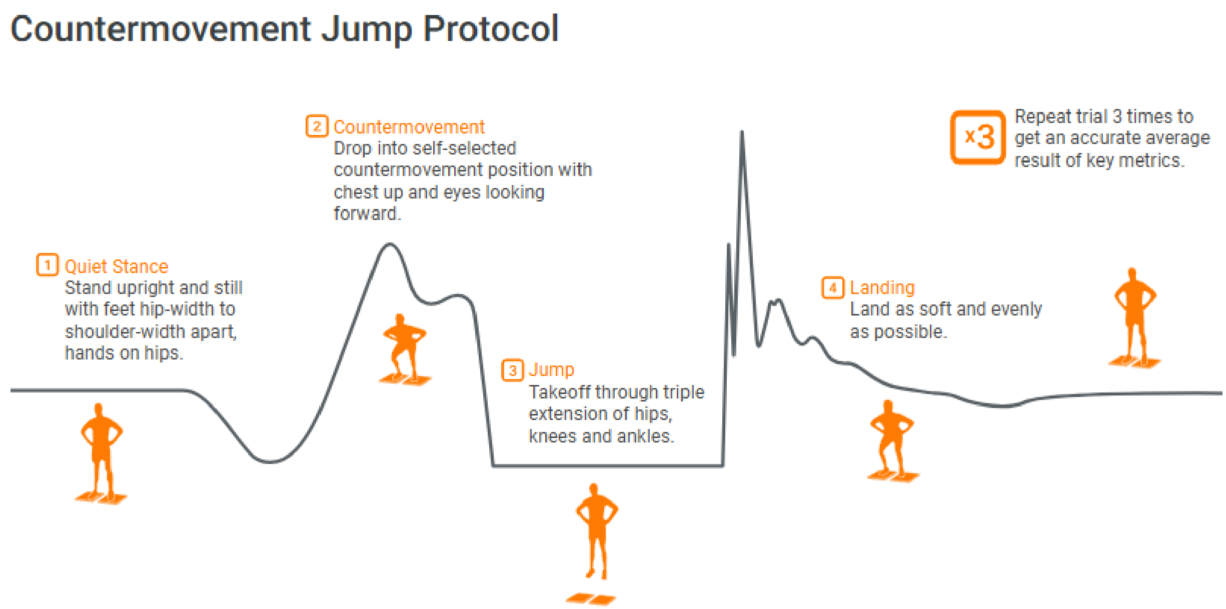

Figure 10: Vald Performance countermovement jump figure to explain the steps of the jump.

Another project that I would like to take on for the team would be lineup analyses. Certain players and certain lineup configurations are likely to perform better than the sum of the players in the lineup, and determining which combinations work best would be helpful in putting the best product on the field. For example, maybe our backup right guard has lower individual blocking stats than the starter, but if the unit as a whole performs better with the backup, we may want to use that configuration more often. This example can be further broken down into situational football. Maybe the backup right guard is better in goal line situations or better on first down? How do we perform in 11 personnel or 12 personnel? Do different wideout combinations work better than others? What about different defensive packages? Should we always go to a nickel package when they have three wideouts, or does our third linebacker outperform our slot corner for certain receiver types? There are unending possibilities for analytics in sports and utilizing them appropriately can help the team improve along the edges to increase the probability of winning each game.

References

Baumer, S. Benjamin, Gregory J. Matthews, and Quang Nguyen. 2023. "Big ideas in sports analytics and statistical tools for their investigation." *WIREs Computational Statistics.* November/December 2023. Vol 15(6): 1-24.

Elmore, Ryan, and Andrew Urbaczewski. 2025. *Introduction to Sports Analytics Using R.* Denver: Prospect Press.

Vald Performance. 2024. "Understanding the Countermovement Jump." Published Mar 11, 2024. https://valdperformance.com/news/understanding-the-countermovement-jump.