Predicting Success of NFL Playcalling

Team 200: Deven Chatterjea, Raghu Dantuluri, Mateen Jangda, James Martin, & Yu Qin

Introduction and Problem Definition

The National Football League, NFL, is the most watched sport in America, with 41% of adults in the US claiming it is their favorite sport to watch [1]. Sports analysts, bettors, and fans all tune in from September to February to watch and during this time, billions of dollars are poured into the sport whether it be from advertising or to gambling.

Due to the popularity of the sport, our group wanted to develop a dashboard/visualization to determine how effective a certain play will be based on multiple game factors like quarter, down, yards to go, and time remaining. In this case, a play in the game of American Football can be defined as a strategy or plan to move the ball down the field. Key definitions and terminologies used in this study, including 'down' and 'play' are outlined in Appendix A.

Our primary focus was to determine how the offensive team's formation, along with other factors, in each situation is expected to perform. Our group highlighted the variable formation, as most current sites that calculate win probability do not take that into consideration. By incorporating play formations, we expect to improve the accuracy in forecasting play success rates.

Literature Survey

Sports analytics, especially involving win probabilities/success rate, has been around since the 20th century, but has recently made big advancements in the 21st century. Our team has reviewed 15 papers related to playcalling and success in the NFL to advance our project.

Goyal [2] and Alamar, B.C [3] both focused specifically on playcalling. Goyal determined that lower predictability playcalling was more successful, while Alamar focused on Yards to Go, to a first down, to determine whether a pass or run play was the better choice. Alamar also examined Expected Points depending on the starting field position. Our team plans to integrate expected points added as one predictor of success. Fernandes, et. al. [4] used the same dataset, NFLsavant.com, as our group. Their study focused on what formation and what type of play will be called and the success rate of those predictions. The classification aspect is something our team uses, but we look to expand to more than just predicting the play called.

Pelechrinis and Papalexakis [5] used statistical bootstrapping to develop a model predicting the outcome of a game. Due to the limited number of games an NFL season has, in our case we analyzed over 2,800 games, we look to employ this method. Gifford and Bayrak [6] took a different approach, using logistic regression and decision trees to predict season standings. This model serves as a baseline for our project, as it provides a foundation for using decision trees and logistic regression to assess drive success. Allen [7] examined fourth-down decisions with logistic regression and Monte Carlo methods; we expand on this by analyzing decision-making across all downs.

Machine learning has been effectively applied to football analysis in studies like that by David et al. [8], who used Artificial Neural Networks (ANNs) to predict game outcomes. While their results are promising for overall success prediction, their model's simplicity may lack what's needed for play-by-play analysis. Lock and Nettleton [9] applied random forest models to determine key features (score, point spread, time remaining, and yard line) are crucial predictors of success (win probability). Their study highlights the importance of feature analysis.

Yurko et al. [10] explored continuous time events to evaluate how play success probabilities evolve in real-time. Although this approach offers valuable insights, our model and visualizations will use discrete events due to our choice of data. Horn et al. [11] attempted a spatial approach by integrating player position data from RFID chips to predict upcoming plays. Although limited success, predicting the next play is viable and could be explored further.

Heiny and Blevins [12] focused on one team and used discriminant analysis on unique variables like weather and relative score. This is something we would consider adding for future studies. Jordan et al. [13] applied game theory, a less mathematical approach, to determining pass or run plays, while Emara et al. [14] found teams switch play types more frequently than random chance, reflecting negative serial correlation. Stilling and Critchfield [15] applied general matching law to study situational play-calling bias. This provided our team with a good insight into the topic but provided little data analysis. Rockerbie [16] modeled optimal play-calling balance, noting a general preference for passing, something our team considered prior to modeling.

Data Preparation and Analysis

A majority of the current "state of the art" success calculators are displayed to the user without allowing users to input their own data. Everything is already predetermined by the ongoing plays. Our group wanted to allow users to input hypothetical scenarios with our website.

The first step of the data collection process involved determining the most applicable source. During the literature review, Fernandes et. Al. used the source our group settled on. NFLsavant.com contained yearly files that contained every single play called during that years' NFL season. Since no scraping was required as the owner had already compiled this data, it was relatively clean and easy to use. The years were then merged into a single source, any rows with missing data were removed, plays were sorted chronologically, and then play numbers were assigned. The plays then had to be organized into drives to calculate the success of individual drives.

Once the data was prepared, the first thing performed was a K-means clustering model, an unsupervised learning method. The resulting clusters represented distinct scenarios, which were summarized by their average feature values and associated outcomes. Additionally, cluster sizes were computed to evaluate the frequency of different play scenarios. A function was then created to compute thresholds for determining play success. The function clustered plays based on their down and identified the proportion of plays in the Red Zone that met predefined success conditions. These thresholds were later incorporated into the supervised learning framework for classification. The thresholds created by our K-means model can be seen in the table below.

Down	Threshold (Yards gained/yards to first)
1	0.54
2	0.62
3	1
4	1

Table 1: K-Means Model results

For supervised learning, the dataset was augmented to classify plays as either successful or unsuccessful. A play was defined as successful if it: resulted in a first down or resulted in a scoring play.

Categorical features, such as team names, formations, and rush directions, were converted into binary variables. A Random Forest Classifier was then trained to predict play success. Bayesian optimization was utilized for hyperparameter tuning, focusing on maximizing the ROC-AUC score. The dataset was split into training (80%) and test (20%) sets, and feature scaling was applied to improve model performance.

The model was evaluated on the test dataset using accuracy, ROC-AUC, and a confusion matrix. Sensitivity and specificity were calculated to assess the model's ability to predict successful and unsuccessful plays. Feature importance analysis was conducted to identify the most influential predictors of play success. The model's performance was further analyzed over time and across different downs. Metrics such as success rates and true positive rates (TPR) were computed for plays grouped by season and down. Visualizations, including bar charts and line plots, were generated to illustrate trends in play success and prediction accuracy over the years. The overall play success found in our data was roughly 30% of play call were successful. Our model was able to accurately predict the outcome of 72.46% of plays, with a TPR of 37% and a TNR of 88%. As can be seen from the model, struggled accurately label positive plays, with a high number of false positives.

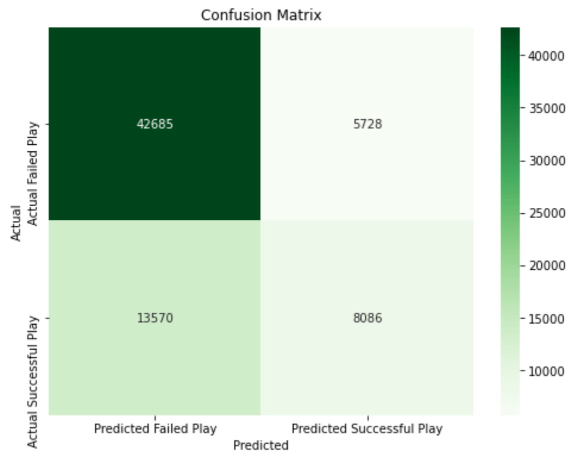

Figure 1: Confusion Matrix of Play Success

To evaluate the model under specific game conditions, the dataset was restricted to plays from the 2020 season on second and third downs. The model's accuracy and confusion matrix were analyzed to validate its predictive capability under these scenarios.

A feature importance chart was created to understand which features are the best predictors of a successful play. The features, Yard Line, Second, Yards to Go, and Minutes were very strong predictors of success. Further insights can be gathered from this to provide to coaches and teams.

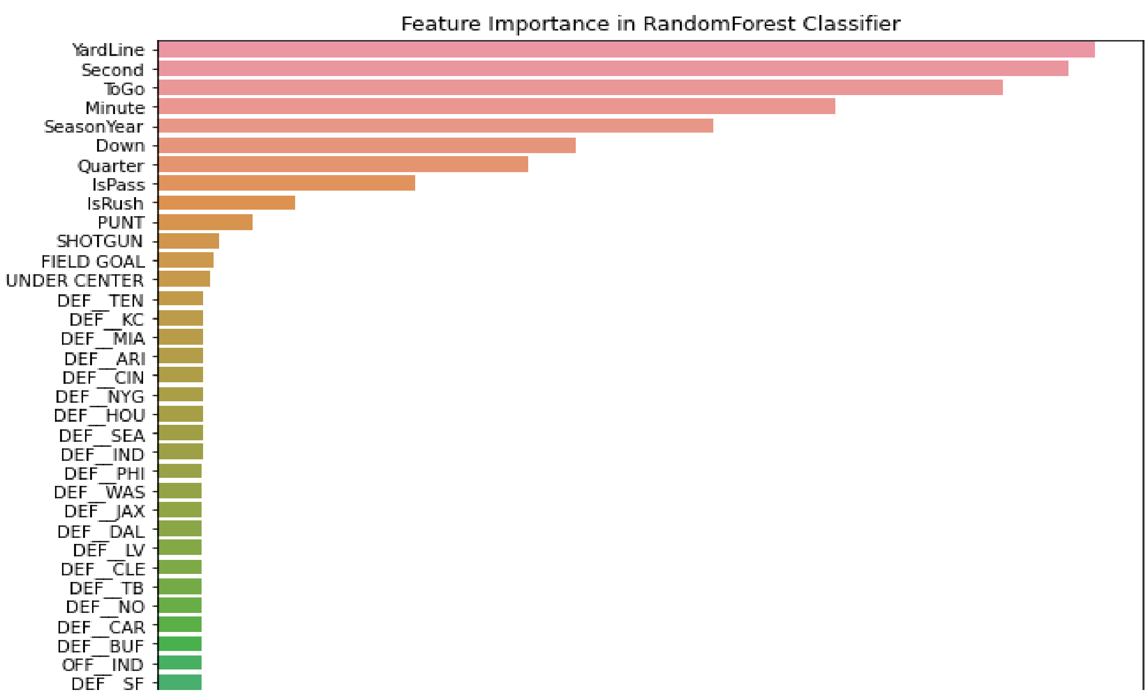

Figure 2: Random Forest feature importance

Experiments and Evaluation

Once the modeling was performed, our team determined the visualization aspect. For this project, we used Python Flask for the website and how to display our results. An interactive webpage displays an animated version of a football field along with a drop-down table of feature selections. The table includes variables such as Quarter, Teams, Formation, and Yards to Go. The football field shows both the current yard marker and the location of the first down marker. At the top, it also provides the success probability. The user is to select all the required inputs, hit “Run Analysis”, and the field and success probability will populate.

The Flask application runs from a pre-trained Random Forest classifier that was developed in the coding section. The figures below show both the on-field representation and the table input page.

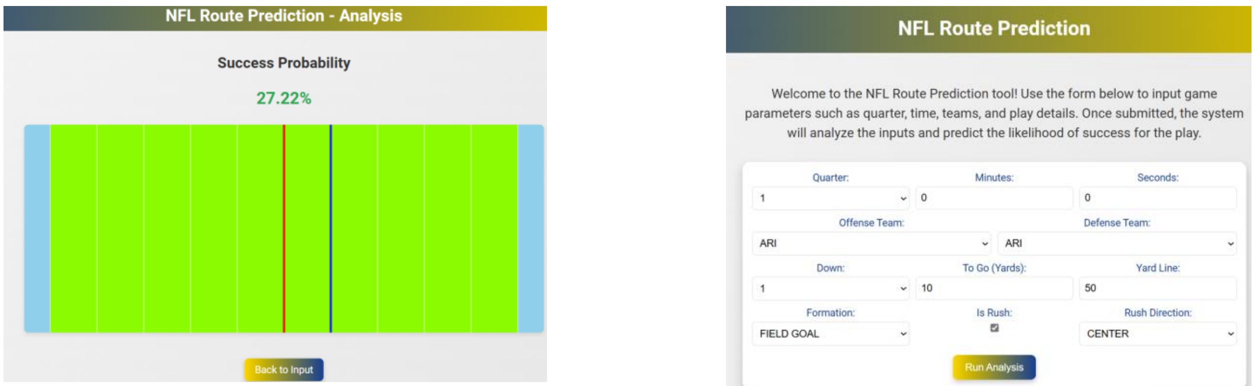

Figure 3: Python flask application interfaces

Conclusion and Discussion

The main goal of this project was to help better analyze success predictions of NFL playcalling and to an extent, that has been achieved. An easy-to-use dashboard was established and can be used by anyone interested in seeing how various features can affect play or drive success.

To further improve this project, a few things could be done. Regarding the dashboard, it is relatively simplistic and certain features, like team logos or visualizations of a run or pass, could be incorporated to better display the play. For the data, some features like weather, rosters, or coaches could enhance the accuracy of the model and visualization. One last big improvement that could be made would be to link the website with real-time data. This would allow a continuous update of how successful the current drive would be. It would be similar to ESPN’s win probability meter.

For this project, Deven, James, and Mateen contributed more to both the report and the presentation.

Appendix A: Definitions

- **Play**: A single action or sequence of actions that starts with the snap of the ball and ends when the ball is dead (e.g., the ball carrier is tackled, the ball goes out of bounds, or an incomplete pass occurs).
- **Down**: A unit of play in which the offense attempts to advance the ball. The offense gets four downs to move the ball 10 yards. If successful, they earn a new set of downs. If not, possession switches to the opposing team.
- **Drive:** A series of offensive plays by one team until one of the following occurs: a score, a turnover, or a turnover on downs.
- **Quarter**: One of the four periods into which a football game is divided. Each quarter is 15 minutes long, with a halftime break after the second quarter.
- **Formation**: The specific arrangement of players on the field before the start of a play. Formations are used to strategize and can indicate whether a team intends to run or pass the ball.
- **Shotgun**: An offensive formation where the quarterback lines up several yards behind the center, rather than directly behind the center.
- **Under Center**: A formation where the quarterback lines up directly behind the center, who snaps the ball directly into the quarterback's hands.
- **No Huddle**: A strategy in which the offense quickly lines up for the next play without taking time to huddle and discuss the play.
- **Field Goal**: A scoring play where the ball is kicked through the opponent's goalposts. A field goal is worth 3 points. Usually occurs on 4th down and within 60 yards.
- **Punt**: A kick made by dropping the ball and kicking it before it hits the ground. Punts are typically used on fourth down when the offense decides to give up possession in exchange for pushing the opposing team farther back on the field.
- **Pass**: A forward throw of the football, usually from the Quarterback to another eligible player. A completed pass occurs when the receiver catches the ball.
- **Rush**: An attempt by a player, typically a running back, to advance the ball by carrying it rather than passing.

- **Interception**: A play where a defensive player catches a pass intended for an offensive receiver, gaining possession of the ball for their team.
- **Red Zone**: The area of the field between the opponent's 20-yard line and the goal line. It is called the "red zone" because teams are in a strong position to score.
- **Fumble**: A loss of possession of the ball by the ball carrier before being tackled or crossing the goal line. The opposing team can recover the fumble to gain possession.
- **Touchdown**: A scoring play worth 6 points, achieved when a player carries the ball into the opponent's end zone or catches a pass in the end zone.
- **Incompletion**: A pass attempt that is not caught by any player, resulting in the play/down ending with the ball returning to the previous line of scrimmage for the next down.
- **Scramble:** When a quarterback, initially intending to pass, is forced to run due to pressure from the defense. Scrambling occurs when the quarterback leaves the pocket (the protected area formed by the offensive line) to avoid being sacked and attempts to gain yards by running or extending the play.
- **Timeout**: A pause in the game requested by a team to stop the clock and discuss strategy. Each team is typically allowed three timeouts per half, and timeouts are often used to conserve time or make tactical adjustments.
- **Kickoff**: The play that begins each half and restarts the game after a scoring play. The ball is kicked by the kicking team from their own 35-yard line to the opposing team.

Appendix B: References

[1] Kerr-Dineen, Luke. “Are “Win Probabilities” Useless? ESPN’s Director of Sports Analytics Explains Why They’re Not.” For the Win, 22 Feb. 2017, ftw.usatoday.com/2017/02/super-bowl-espnwinprobability-atlanta-falcons-new-england-patriots-stats-tom-brady. Accessed 12 Oct. 2024.

[2] Goyal, U. (2020). Leveraging Machine Learning to Predict Playcalling Tendencies in the NFL (dissertation). Leveraging machine learning to predict playcalling tendencies in the NFL.

[3] Alamar, B. C. (2010). Measuring Risk in NFL Playcalling. Journal of Quantitative Analysis in Sports, 6(2). https://doi.org/10.2202/1559-0410.1235

[4] Joash Fernandes, C., Yakubov, R., Li, Y., Prasad, A. K., & Chan, T. C. Y. (2020). Predicting plays in the National Football League. Journal of Sports Analytics, 6(1), 35–43. https://doi.org/10.3233/JSA190348

[5] Pelechrinis, K., & Papalexakis, E. (2016). The Anatomy of American Football: Evidence from 7 Years of NFL Game Data. PLOS ONE, 11(12), e0168716. https://doi.org/10.1371/journal.pone.0168716

[6] Gifford, M., & Tuncay Bayrak. (2023). A predictive analytics model for forecasting outcomes in the National Football League games using decision tree and logistic regression. Decision Analytics Journal, 8, 100296–100296. https://doi.org/10.1016/j.dajour.2023.100296

[7] Keith, A. (2021). Bridging the Paradox: Using Markov Chains to Supplement National Football League Play-Calling Strategy [Review of Bridging the Paradox: Using Markov Chains to Supplement National Football League Play-Calling Strategy]

[8] David, John A., Pasteur, R. Drew, Ahmad, M. Saif and Janning, Michael C.. "NFL Prediction using Committees of Artificial Neural Networks" Journal of Quantitative Analysis in Sports, vol. 7, no. 2, 2011. https://doi.org/10.2202/1559-0410.1327

[9] Lock, Dennis and Nettleton, Dan. "Using random forests to estimate win probability before each play of an NFL game" Journal of Quantitative Analysis in Sports, vol. 10, no. 2, 2014, pp. 197-205. https://doi.org/10.1515/jqas-2013-0100

[10] Yurko, R., Matano, F., Richardson, L. F., Granered, N., Pospisil, T., Pelechrinis, K., & Ventura, S. L. (2020). Going deep: models for continuous-time within-play valuation of game outcomes in American football with tracking data. Journal of Quantitative Analysis in Sports, 16(2), 163–182. https://doi.org/10.1515/jqas-2019-0056

[11] Horn, Hayley; Laigaie, Eric; Lopez, Alexander; and Reddy, Shravan (2023) "Using Geographic Information to Explore Player-Specific Movement and its Effects on Play Success in the NFL," SMU Data Science Review: Vol. 7: No. 2, Article 3.

[12] Heiny, Erik L. and Blevins, David (2011) "Predicting the Atlanta Falcons Play-Calling Using Discriminant Analysis," Journal of Quantitative Analysis in Sports: Vol. 7: Iss. 3, Article 2

[13] Jordan, J. D., Melouk, S. H., & Perry, M. B. (2009). Optimizing Football Game Play Calling. Journal of Quantitative Analysis in Sports, 5(2). https://doi.org/10.2202/1559-0410.1176

[14] Emara, N., Owens, D., Smith, J., & Wilmer, L. (2017). Serial correlation in National Football League play calling and its effects on outcomes. Journal of Behavioral and Experimental Economics, 69, 125– 132. https://doi.org/10.1016/j.socec.2017.01.007

[15] Stilling, S.T. and Critchfield, T.S. (2010), THE MATCHING RELATION AND SITUATIONSPECIFIC BIAS MODULATION IN PROFESSIONAL FOOTBALL PLAY SELECTION. Journal of the Experimental Analysis of Behavior, 93: 435-454. https://doi.org/10.1901/jeab.2010.93-435

[16] Rockerbie, D. W. (2008). The Passing Premium Puzzle Revisited. Journal of Quantitative Analysis in Sports, 4(2). https://doi.org/10.2202/1559-0410.1093

[17] https://nflsavant.com/

[18] https://www.pro-football-reference.com