**Big Data White Paper**

ECON 695

Alec Stuedemann, Jason Gong, Sowmya Subramaniam

## 

## 

## 

## 

## **Abstract:**

This paper analyzes decision-making in NFL fourth-down plays using a comprehensive dataset from 2021 to 2023. Employing advanced machine learning models, the study identifies key strategic and environmental factors that influence critical game decisions. Through logistic regression and feature importance analysis, we pinpoint predictors of successful fourth-down conversions, offering actionable insights for NFL teams. The research not only elucidates the complex dynamics of play-calling but also demonstrates the significant impact of contextual factors such as time dynamics and player positioning on the outcomes. These insights empower teams to optimize their play strategies in high-pressure situations, enhancing their overall competitive edge in the league.

## Introduction:

In the dynamic environment of NFL games, fourth-down decisions are critical and can significantly affect game outcomes. These decisions - whether to attempt a conversion, punt, or kick a field goal - are influenced by multiple factors including game context, player performance, and environmental conditions. This research utilizes a robust dataset, statistical tests, and machine learning methods to explore these factors, extract meaningful patterns, and offer strategic insights through the lens of big data analytics.

## Methodology:

This study utilized an extensive dataset from the NFLverse GitHub repository, encompassing every play from NFL games dating back to 1999, with a focused analysis on the 2021 to 2023 seasons. The year 2020 was excluded due to the atypical disruptions caused by the COVID-19 pandemic, which could potentially skew the predictive modeling and analysis. Fourth-down plays within NFL games represent critical moments where teams must decide whether to attempt to advance the ball and secure a first down, punt, or try for a field goal, depending on their field position. These decisions are pivotal to game outcomes and provide a rich context for studying decision-making under pressure, a key interest for players, general managers, and coaches who are constantly seeking competitive advantages. The methodology involved preprocessing the data using Python's pandas library to handle tasks such as data cleaning, normalization, and the handling of missing values. Features considered in the analysis included play type, player formation, game state variables, and other contextual factors impacting fourth-down decisions. To analyze the factors influencing the outcomes of fourth-down plays, logistic regression models were developed using Python's sklearn library. This approach allowed for the examination of the impact of various strategic and environmental factors on the likelihood of a successful fourth-down conversion. The dataset, comprising over two million observations, provided a robust basis for statistical analysis and model validation. Overall, the study aimed to discern the critical factors influencing fourth-down decision-making in NFL games, thereby providing insights that could help teams enhance their strategic approaches during these pivotal game moments.

## Analysis:

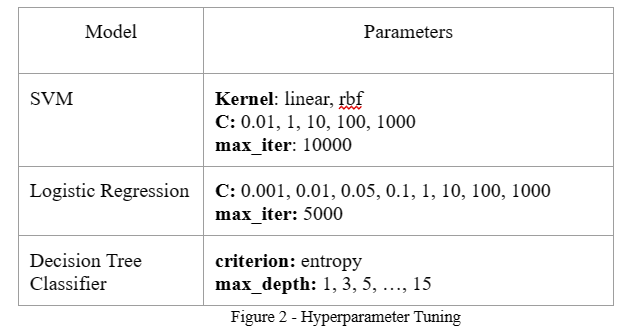
The initial phase of the analysis involved identifying and selecting the most impactful variables from the extensive dataset. Key variables were directly available, but to enhance the model's effectiveness, additional data manipulation was necessary to derive new variables. A significant variable hypothesized to influence fourth-down decisions was whether a timeout was called before the play. This situation in a game when a timeout is called before a fourth down can be critical in determining the outcome of the play. When the defense calls a timeout, they can change their formation, or the offense can change the play, affecting the outcome of the play. This variable was not directly present in the dataset, necessitating a derived approach using existing columns. Specifically, the analysis involved filtering instances where the "fourth\_down\_converted" and "fourth\_down\_failed" fields were not null. This was paired with conditions where 'time\_out' equaled one and "play\_type" was listed as 'no play.' The data was further refined to identify rows where either "fourth\_down\_converted" was not equal to "fourth\_down\_failed" or "timeout" was 1 for 'no play' entries. This process helped derive a new binary variable, "time\_out\_before," which indicates whether a timeout was called immediately before each fourth-down play. This variable was hypothesized to be a strategic factor influencing the play's outcome, encapsulating the strategic use of timeouts in critical game situations. After this, they narrowed down the data to only include 4th down plays, as that was their focus. Since "play\_type" was a categorical variable, one-hot encoding was necessary. This process replaced the original values with their encoded representations and added a new column "is\_pass" based on the one-hot encoding of "play\_type”.

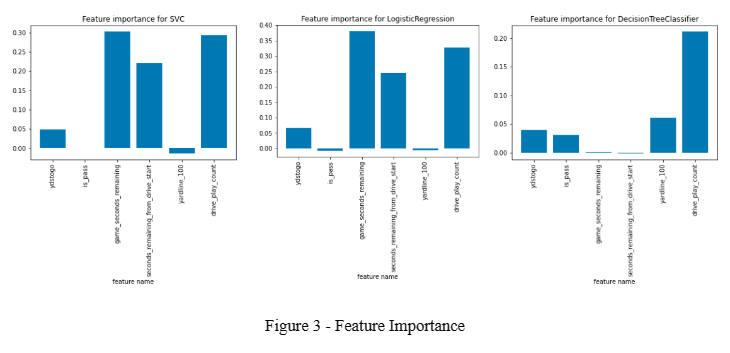
Data imputation was also required for the variables wind and temp. For missing values or NaNs in the temp variable, they replaced them with a temperature of 70, and for wind, with 0. These were considered sufficient average values for replacement. This variable captures the number of seconds remaining in the game from the start of each team's drive throughout the game. This was achieved by determining whether the drive started before or after the 4th quarter and computing the game clock to determine the seconds remaining from the start of the drive. Finally, they assembled the set of predictors, including "ydstogo", "is\_pass", "game\_seconds\_remaining", "seconds\_remaining\_from\_drive\_start", "score\_differential", "yardline\_100", "play\_clock", "temp", "wind", "shotgun", "drive\_play\_count", and "timeout\_before". The dependent variable then became “fourth\_down\_converted”.

The models chosen to test were support vector machine, logistic regression, and decision tree classifier. These models are best suited to binary classification and fit the context of the problem. Of the models chosen, the decision tree classifier has the best explainability but is prone to overfitting. The other two can be good for avoiding overfitting due to regularization terms but have lower explainability with high-dimension data. The choice of kernel function significantly influences the flexibility and complexity of the decision boundary learned by the SVM model. This analysis varied the regularization parameter C across a range of values (0.01, 1, 10, 100, 1000) to control the trade-off between maximizing the margin and minimizing the classification error. With this wide range of C values the result that may narrow the margin and induce overfitting of the data was controlled for. To control for convergence, the max number of iterations was set to 10000. Similarly the range of regularization terms in the logistic regression model help control for overfitting. Additionally, the SVM struggles to predict data that is not separable if the proper kernel is not used.

An 80%-10%-10% train-validate-test split of the data was used for model selection and hyperparameter tuning. Grid search was utilized to find the optimal model and hyperparameters. See Figure 2 for the hyperparameters tried for each model. The SVM and logistic regression models needed a max iteration parameter because the models failed to converge. Scaling the data did not solve this deficiency.

A second grid search with several features removed did not significantly impact the model to reduce overfitting. The final features utilized by the models were yards to go, pass or run, seconds remaining in the game, seconds remaining at the start of the drive, yard line, and number of plays in the drive. See Figure 3.





Several interesting features emerged with varying degrees of importance. After reviewing the initial feature importance visualizations, we opted to eliminate some variables that had minimal impact on reducing overfitting. Ultimately, the final model comprised six variables: "ydstogo," "is\_pass," "game\_seconds\_remaining," "seconds\_from\_drive\_start," "yardline\_100," and "drive\_play\_count."

Initially, it was anticipated that "ydstogo," representing the distance to the first down, would significantly influence the model's performance. However, its impact turned out to be substantial but not as pronounced as expected. Surprisingly, "drive\_play\_count" emerged as the most influential variable. This finding aligns with the logic that if a team demonstrates offensive prowess throughout the drive, there's a higher likelihood of maintaining momentum and opting to go for it on 4th down.

In summary, the model provides valuable insights into the decision-making process of NFL coaches when determining whether to go for it during games. Despite some unexpected findings, the selected variables offer a comprehensive understanding of the factors influencing 4th down decisions.

The logistic regression model with inverse regularization term 0.01 performed the best on the validation data and was selected for the model. See Figure 4 for a comparison of the other models. The small value for the inverse regularization term is likely used to account for larger values in the dataset because scaling was not used. The model had an accuracy score of 0.848 on the validation data and 0.835 on the unseen test data. The similarity of these scores implies that the model did not have a high level of overfitting.

## 

## 

## 

## 

## 

## 

## 

## Results:

The logistic regression model emerged as the most effective, optimized through grid search to determine the optimal model and hyperparameters. It leveraged significant predictors such as "ydstogo", "is\_pass", and game context variables to evaluate their influence on fourth-down conversions. A meticulous feature importance analysis refined the focus to the most impactful predictors, notably "drive\_play\_count", which unexpectedly proved to be the most influential in predicting fourth-down decisions. The finalized model demonstrated robust predictive capabilities with an accuracy score of 84.8% on the validation data and 83.5% on the unseen test data. These results underscore the model's ability to effectively identify strategies that lead to successful fourth-down outcomes. Notably, the model highlighted the significant influence of strategic timeouts and other game situational variables, providing deep insights into the nuanced decision-making process during critical game moments. Overall, these results not only validate the effectiveness of the analytical approach but also enhance understanding of the tactical considerations NFL teams employ during pivotal moments of the game.

## Discussion:

**Addressing Potential Confounders and Mitigation Strategies:**

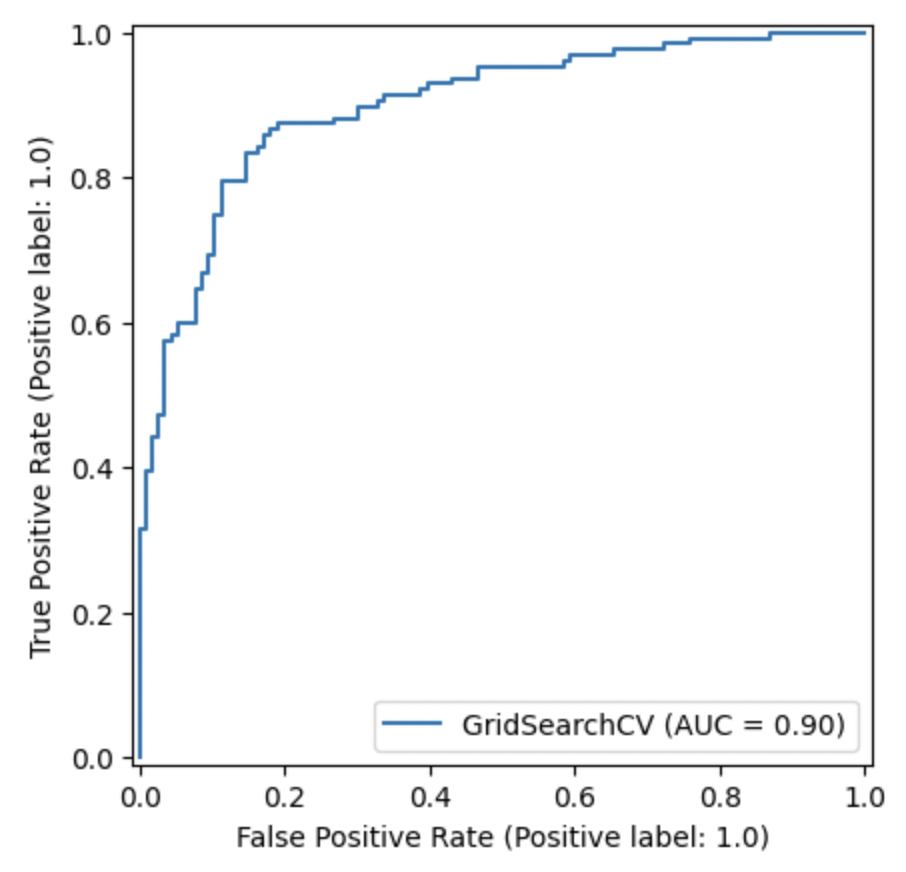
In the process of analyzing fourth-down plays in NFL games, several potential confounders were identified that could skew the perceived effectiveness of various strategies. These include game-specific variables such as weather conditions (wind and temperature) and game location, which can dramatically affect outcomes. For instance, adverse weather might impede passing games or alter field goal decisions. Additionally, variations in player skills, team strategies, and coaching philosophies also play a crucial role, as they may impact the effectiveness of fourth-down plays differently across teams. Moreover, the strength and defensive strategy of the opposing team can further confound results, with stronger defenses possibly making offensive strategies seem less effective than they are.

To mitigate these confounders and enhance the reliability of the findings, a multifaceted approach was employed. Firstly, data cleaning and normalization were meticulously carried out, particularly for environmental variables. Missing data for elements like wind and temperature were imputed using historical averages to maintain dataset integrity and reduce bias. Secondly, through feature engineering, new variables such as 'timeout before the play' were created. This helped to isolate the impacts of strategic decisions from other influencing factors, providing clearer insights into the direct effects of timeouts on play outcomes. Further, statistical controls were implemented by using logistic regression models that included confounding variables as covariates. This approach effectively isolated specific effects, ensuring that the impacts observed could be directly attributed to the strategies under examination rather than extraneous influences. Additionally, model validation was rigorously performed using a robust train-validate-test split, coupled with Grid Search for hyperparameter optimization to confirm that the model's predictions were both reliable and not overly fitted to the dataset's idiosyncrasies. Considering the complex nature of causal relationships within the data, advanced analytical techniques such as propensity score matching or instrumental variable analysis were also considered. These methods are particularly valuable for addressing confounding when exploring the causal impacts of specific strategies on game outcomes.

In summary, by comprehensively addressing potential confounders and rigorously testing mitigation strategies, this analysis not only enhances the credibility of the findings but also ensures that the insights provided can effectively inform strategic decisions in real-world NFL scenarios.

## Conclusion:

The comprehensive analysis of NFL fourth-down play decisions underscores the transformative impact of data analytics on enhancing strategic game management and decision-making. By examining over two million play-by-play instances from 2021 to 2023, the



study revealed key insights into the dynamics influencing critical game moments.

The Receiver Operating Characteristic (ROC) curve is a fundamental tool for evaluating binary classification systems, such as those used in our NFL fourth-down play decision analysis. This curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. The TPR, or sensitivity, indicates the proportion of actual positives correctly identified (e.g., successful fourth-down conversions), while the FPR measures the incorrectly classified negatives (e.g., failed conversions marked as successful). An effective classifier’s ROC curve will approach the top-left corner of the plot, demonstrating high sensitivity and low false positive rates, crucial for distinguishing between successful and unsuccessful plays accurately. The area under the ROC curve (AUC) serves as a single measure of the model's overall discriminative ability, with an AUC of 1.0 representing perfect discrimination and 0.5 indicating no discriminative ability. In our project, the logistic regression model showed robust predictive power with an AUC of 0.90, indicating excellent capability in differentiating between outcomes. The ROC curve is invaluable for optimizing model thresholds and enhancing the decision-making process in high-stakes game situations, ensuring that strategic decisions are both effective and efficient. The logistic regression model, which exhibited a robust predictive power with an AUC of 0.90, highlighted the effectiveness of the analytical approach in discerning between successful and unsuccessful fourth-down conversions.

The findings challenge traditional conservative play-calling norms by demonstrating that variables such as game clock time and score differentials are crucial in decision-making processes. It was found that plays earlier in the drive and closer to midfield tend to have higher success rates. This insight not only challenges the norm but also integrates key economic principles like opportunity cost and resource optimization, providing teams with a strategic edge that can enhance performance and economic returns through improved game management and increased fan engagement.

To ensure the integrity and reliability of the findings, stringent methodological safeguards were implemented. These included addressing potential confounders, such as weather conditions and player performance, by incorporating these elements as control variables in the models. This allowed us to isolate the effect of strategic decisions from external influences effectively. Additionally, commitment to data integrity was highlighted by exclusion of data from the atypical 2020 season and rigorous data cleaning, feature engineering, and validation processes.

Looking forward, there are substantial opportunities for further refining the models with emerging data and advancing analytical techniques. This continual enhancement will likely improve prediction accuracy and uncover new influential factors. It is recommended that teams incorporate these insights into their real-time decision-making frameworks to optimize fourth-down strategies and overall game tactics. Moreover, by extending this analytical approach to other areas of game management, there is potential to significantly impact economic aspects of sports management, such as ticket sales, merchandise, and media rights.

In conclusion, this research not only advances understanding of strategic decision-making in high-stakes NFL scenarios but also sets a precedent for the broad application of sophisticated data analytics in sports. As NFL teams continue to embrace and refine these techniques, they can anticipate not only more successful fourth-down conversions but also a more strategic and economically sound game plan. This study serves as a testament to the powerful role of data analytics in sports, proving that with the right data and analytical tools, teams can significantly enhance their decision-making processes and overall performance.