Best Plays to Run in NFL during any given Circumstance through Every Play in NFL 2010-2020.

Charles Mackey

Johns Hopkins Whiting School of Engineering  
cmackey5@jh.edu

**Abstract**

The National Football League (NFL) is a highly competitive where each decision and play execution need to be in sync at all times. Having a playbook; that involves a variety of players, offensive schemes, take into account defensive schemes, and the probability of execution is crucial to the game. This research paper will examine the best plays to run during a given circumstance in the NFL. This will include the decision process based on the current situation, and what plans would be most effective. By analyzing the best options that coaches and players have during any circumstance this paper is designed to provide coaches, players and fans with a better understanding of the factors of success on the field.

Introduction

The NFL is the most popular sports league in the United States, and it’s played at the highest level in the world. Games are won and lost based on the effectiveness of plays, and by teams that can analyze their opponents’ strengths and weaknesses. From there coaches need to adapt their teams play accordingly to have a better chance of winning. Creating an AI-enabled system will help teams analyze the effectiveness of plays by using machine learning algorithms that can provide real-time valuable insights that can optimize a team’s performance.

Right now, NFL’s Next Gen Stats [1] with the use of Amazon Web Services (AWS) is capable of capturing the information needed to make decisions, but not all of this data is not utilized in real-time. Currently all the stats collected with Next Gen Stats are after the fact and do not use any Artificial Intelligence (AI) that we know of. Fancy and elaborate graphics of wide receiver (WR) routes, running plays, and defensive schemes are demonstrated on the tv screen, but aren’t relevant in a game-time setting where every second matters. Play calling by coaches, and the time it takes to make a decision on the field are hindered by several factors.

A coach’s biased that might overlook a specific offensive or defensive schemes, deciding the most efficient personnel on the field, and taking player’s personal abilities into account within forty, twenty-five, or ten seconds to make a crucial decision is one of the most challenging aspects of the NFL and all sports in general [2].

With open datasets from Kaggle[3], and Next Gen Stats every play of every game and the outcome of each play is captured and can assist in making real-time decisions.

In this research paper the type of analysis, systems used, database structure, and preprocessing necessary for an AI-enabled system to work for NFL play calling is outlined.

**Key Partnerships**

For this system to work it would need to focus on NFL teams and their coaching staff. The average coaching staff in the NFL consists of around 10 members. Each coach usually has a different role, but for the most part they’re relying on; existing game film for their team and the opposition (which they might not remember on the spot), analytics and scouting reports on the opposition (again, very difficult to remember on the spot), headsets for communication from the field to the booth (most play calling comes from the booth), and their extensive playbooks (which can be biased based on a coach’s personal preference, the score, or the opponent).

All play calling decisions have a chain of command based on the type of play and current situation. Usually, a coordinator will communicate play options to the head coach and the head coach will relay that information to the quarterback. This might involve changing the players on the field, and then the quarterback needs to relay this information to players in a huddle. Previous time is lost depending on the game situation and time left in the game.

**Key Activities**

The AI-enabled system for NFL play analysis will use machine learning algorithms to analyze and predict the effectiveness of plays.

1. The system will be trained on historical play data, which includes game situations, player statistics, and play outcomes. The system will be able to recognize patterns in the data and identify which plays are most effective based on the situation.
2. Data Analysis: once the data is collected it must be analyzed to identify patterns and insights.
3. Model Development: Once the analysis has been done a predictive model will be created that can be used to make real-time play calling decisions during games. This may involve using a combination of statistical models and machine learning algorithms.
4. Integration with Game Systems: right now, in the NFL communication between multiple coaches and then to players on the field is one of the major setbacks. There needs to be an interface where this information can be relayed more efficiently.
5. Testing, refinement, and deployment: this portion of creating an AI-enabled system for play calling will be the most difficult. Again, coaches have their own personal preferences or bias in play calling that will take time for them to accept, adopt, and trust a system that overrides their initial instinct during a game-time decision. Overall, this system would need to be maintained and monitored over time to see if there is an improvement in performance by the team or if any updates need to be made.

**Key Resources**

This AI-enabled system will be built using a combination of open-source tools and custom-built software. The system will be based on the Python programming language, and the following libraries will be used to implement the system.

1. Scikit – for data preprocessing, feature selecting, and model evaluation.
2. Numpy – for data manipulation and analysis
3. Pandas – for data manipulation and analysis

Asides from the software necessary teams and coaches would need to be willing to provide the necessary data that might not be available to the public. This would involve playbooks (sometimes involve secret plays), actual player’s health, and the willingness to let decisions be made by the system.

Open-source data from Kaggle that has been preprocessed and has gone through the correct model evaluation will ultimately assist in play calling.

**Database Structure**

In terms of the database structure there should be four main aspects that need to be stored.

1. Play: information about each play, the down, the distance, and the play type.
2. Game: information about the date, location, and final scores or previous games.
3. Player: information about each player involved including their position, team, and statistics
4. Opponent: Whether on offence or defense, knowing the opponent’s tendencies will help determine weaknesses and play calls can be made to exploit those weaknesses.

Overall, a database structure that can handle large amounts of data, be flexible to handle changes during each game, and perform quick queries to provide insight would be necessary.

**Preprocessing Data:**

There would need to be multiple steps in preprocessing the data for this AI-enabled system to work.

1. Data Cleaning – Removing irrelevant or duplicate data points from the dataset. This could include several moments in the game that are usually irrelevant. Timeouts or intentional ‘player injuries’ that disrupt the game.
2. Feature Selection – Probably one of the most important steps in preprocessing to identify the most relevant features in determining the effectiveness of plays.
3. Feature Engineering – This would involve creating new features by combining the features selected to improve the accuracy of the model.

**Methods Used**

Ensemble learning, regression, and reinforcement learning will all be used in creating this AI-enabled system for play calling in the NFL.

1. **Ensemble Learning** – This involves combining the predictions of multiple machine learning models to produce a more accurate prediction on the effectiveness of play calling in the NFL. Using this learning approach would be used to combine multiple modes that are trained on different types of data. Player performance, team performance, historical game data can all be used to generate more accurate play recommendations.
2. **Regression** – With regression supervised learning can help predict a continuous output variable based on one or more input variables. Regression could be used to predict the expected outcome of a play based on the features selected. This may include the players involved, the opponent’s defensive scheme or the specific number of yards needed.
3. **Reinforcement Learning** – Is a machine learning technique that involves training an agent to take actions in an environment to grant or maximize a reward signal. Reinforcement learning could be used to train an AI agent to make play calling decisions that maximize a team’s chances of getting a first down, scoring points, or winning the game. With reinforcement learning using a simulated game situation or expected decisions that are normally made by human coaches could result in effective and efficient play calls.

Multi-Attribute Tradespace Exploration/Search (MATE)

Although I don’t think this has been done before in the NFL, MATE is a decision-making approach that involves evaluating a range of possible options based on multiple criteria or attributes, such as the likelihood of success, the potential payoff, and the risk involved.

The attributes or features that we would like to optimize would need to be defined, and a list of possible plays would be determined.

This approach could be useful for considering a wide range of possible play calls and selecting the one that is most likely to lead to a favorable outcome based on various factors.

With this method instead of only considering an offensive or defensive scheme or players on the field it always for a trade-off between different play calls in game situations.

Monte Carlo Tree Search (MACTS) [5]

MACTS is a search-based approach that uses Monte Carlo simulations to evaluate different possible decision paths and identify the optimal course of action based on a set of predefined goals or objectives [5].

This approach will be useful for identifying the best play call based on a variety of factors, such as the likelihood of success, the potential payoff, and the risk involved, while also taking into account the uncertainty and randomness that can happen on any given play. (Injury, interception, deflated balls, or poor refereeing.) Monte Carlo method is.

**How we are different:**

While predictive analytics are currently used to determine the likelihood of success on any given play using an AI system will provide additional assistance with machine learning and natural language processing (NLP).

The main benefit from what is currently being used is that with the correct system in place teams, players and coaches can all get real-time analysis.

This is especially true when each player and their abilities are taken into account. A wide receiver or running back might have an advantage against the defensive scheme that a coach or player might not immediately recognize.

With an AI-enabled system that uses pattern recognition an immediate insight could be used to exploit those patterns in play calling. It could be determining a wide receiver’s route or a quarterback’s tendencies on how they pass the ball.

Another way that we can be different is with NFL WAR. NFL WAR percentages use logistic regression and we have not decided on that as a solution yet [4].

NFL WAR evaluates plays and players but doesn’t evaluate the best play. Where we don’t care what position a player is in. We’re also not tied to logistic regression as of yet.

We can use cross fold validation as well as reinforcement learning. In terms of data, we have ten years of data and can measure performance and our model by comparing WAR percentages of success with mean square error.

**What approaches have been taken?**

If we’re comparing our system to a coach’s biased, we can take the biased out. We can also compare our success to other teams that aren’t using our system. If we’re close to ground truth then we can measure our success. We can use the system to predict the success of our system against teams that aren’t using our system.

**Conclusion**

An AI-enabled system for NFL play calling analysis will provide valuable real-time insights to teams that will be used to optimize their performance on the field without coach and player play calling bias. This system will use machine learning algorithms to analyze historical play data and help predict the effectiveness of plays in every changing game situation. With preprocessing the data, feature selection and feature engineering an accurate model can be created.

Although AI can provide more personalized, robust, and real-time decision-making It would most likely be used in conjunction with coaches and their play calling.

**References**

1. NFL (n.d.). Next Gen Stats. Next Gen Stats - NFL. <https://operations.nfl.com/gameday/technology/nfl-next-gen-stats/>
2. Rookie Road (n.d.). Football Play Clock. <https://www.rookieroad.com/football/101/play-clock/>
3. Zimmerman, C. (n.d.). NFL Play by Play. <https://www.kaggle.com/datasets/charliezimmerman/nfl-play-data-2010-2020/code>
4. PFF (n.d.). PFF WAR. PFF WAR. <https://www.pff.com/war>
5. Toward Science (n.d.). Monte Carlo Tree Search for NFL. Https://Towardsdatascience.com. <https://towardsdatascience.com/using-monte-carlo-tree-search-for-your-fantasy-football-draft-6509b78a1c20>