Extra Point Decision-Making in the NFL using Reinforcement Learning

Michael W. Dacus, Ian M. Hokaj, and Tyler M. Weiss
Stanford University, Department of Aeronautics and Astronautics, AA 228
496 Lomita Mall
Stanford, California, 94305

Abstract

In the game of American Football, specifically with regards to the National Football League (NFL), proceeding with kicking an extra point (PAT) has been an "all-to-regular" theme, without paying much thought into possibly going for a two point conversion. Due to recent changes and other game time factors, coaches and teams are beginning to rethink strategies for post-touchdown scoring opportunities. This project aims at determining when it is best to go for a two point conversion at any point in the game by the problem as a Markov Decision Process (MDP). Through multiple State Space scenarios, it was found that kicking offers the best short term reward, but will not necessarily win games. Depending on the point spread and number of drives, going for a two point conversion offers a better chance of winning.

1. Introduction

1.1 Historical Background

Extra points - whether attempting a PAT, or completing a two point conversion - has seen the most revisions and change to the field of play in American Football since the early foundations of the sport. With football's origins deriving from Rugby and soccer, football would immediately proceed with kicking an extra point upon a team scoring a touchdown. In the late 1800s, kicking an extra point was worth more than a touchdown (4 points, whereas a touchdown was only 1), but the values were later revised and updated in 1912 to the current scoring values of modern-day football(Brown 2021).

As the game of football evolved (with additional minute changes to the game made) kickers were beginning to successfully make all PATs upon scoring a touchdown (in the NFL, PAT percentages were well into the ninety percentile range), and spectators and analysts alike began to see the PAT as a "free point." In effort to reduce PAT percentages and add additional post-touchdown scoring opportunities, in 1994, the Competition Committee decided to create the two-point conversion, where teams have the option to either run or pass on one play in attempting to score for two additional points (two point conversions had already been added in other conferences, like the USFL and Collegiate Football). Teams began to use the new two-point conversion as a means for getting ahead early or closing the point spread later in the game; however, it still did not make the PAT kick more difficult for NFL kickers. With PAT completion percentages still in the mid-to-high 80s and 90s, the Committee needed to review on how to make PAT difficult. The solution was simple: in 2015, the Committee moved the spot of the PAT kick from the 2 to the 15 yard line, resulting in a 33 yard kick (Tomlinson 2015) By increasing the PAT distance, it would make it more difficult for kickers to successfully make the PAT kick. Since the year the ruling change was made, there has been a 5 percent average decline in PAT kick percentages among the 32 teams. (Smith 2021)

1.2 Review of Coaching Decision-Making

Over the years, coaches have devised a series of decision-making processes that relied on a number of factors, but weighted subjectively based on previous experiences. At the time, Dick Vermeil, head coach of UCLA Football (1974-75) created a chart scheme based on point spread that determined when it would be best to go for two (Alder 2020). Based on the chart, Coach Veirmeil decided to go for two points that would increase point spread beyond a standard touchdown + field goal combination (6), or get within a multiple of a field goal. However, this could be problematic if the team has low two point conversion percentage (since the chart goes more for two points that in reality). Not only conversion percentages, but time of game can also affect coaches' decision-making, which could lead to irrational decisions. Based on data of coaches' decision-making on other parts of the game (onside kicks), coaches are highly risk-averse and loss-averse (conservative play calling, while also thinking they are in a worse state) (Urschel and Zhuang 2011).

1.3 Problem Statement and Objective

From the evolution and rule changes behind the game, to coaching, play calling and gametime factors, there is a necessity to find optimal solutions as whether a team should kick a PAT, or go for a two point conversion. This project aims at determining the best action by formulating the problem as a Markov Decision Process (MDP) based on defining the State, Action, Reward and Transition in three different Scenarios of increasingly accurate game representation, and parameterizing based on team performance history (PAT kick percentage, two point conversion percentage, kick and two point stopping percentage).

2. Problem Setup/MDP Definition

2.1 Scenario 1

Scenario 1 is based on the premise that we are deciding whether to go for the PAT kick or two point conversion at an instantaneous point of the game, without taking into consideration the point spread or the time left in the game. With that in mind, the State Space is rather small and discrete: it consists of scoring a Touchdown, making a PAT kick, making a two point conversion, missing either extra point, and a terminal state. The action space is simply attempting a PAT [kick] or going for two. The Reward Model is based on the action and next state; if the kick was made, then the reward was set to one, and for two point conversions, two. Finally, the Transition Model was defined by parameterizing transitions from the touchdown state to either the extra point made or missing the extra point using a probability mass p (Bernoulli parameter), which is defined as the probability of the team making the extra point and the opponent not being able to stop the ball, shown in Equation 1:

$$p = p_{suc}(1 - p_{stop}) \tag{1}$$

In total, there will be two p_{suc} and two p_{stop} values, (one for each action), which are defined by team probabilities sourced from 2020 NFL statistics. If p represents the probability they will make the extra point, than probability of missing will simply be 1-p. All other states will transition to terminal state (all zeros except 1).

2.2 Scenario 2

Scenario 2 uses the same framework as Scenario 1, with the major change of redefining the State Space as point spread between two teams. If the state value is positive, then the team [attempting the extra point] is ahead; if the value is negative, then the team is behind. The Transition Model is reorganized using the same probability mass p as defined in Equation 1. The Reward Model is a modified version of Scenario 1; for simplicity, actions are rewarded at states where the team has the ability to force the opponent to score an additional touchdown to gain the lead (multiples of 6) or close touchdown gaps. Rewards also disincentivize for PAT kicks when the team has the ability to pass the other team in points and is also behind in scoring. Finally, at any state, an additional reward for going a kick or for two is modeled as an exponential decay function, as shown in Equation 2:

$$R(s, a) = Ce^{\pm s/15} \tag{2}$$

where C=1 (when a=kick) or C=2 (a=score). The power is negative when s>0 because it doesn't matter which action to take, as long as the team is up by multiple touchdowns, the rewards will be almost the same. For the other case, it is positive when s<0 because whichever action the team takes will not guarantee a win (if they are down multiple touchdowns).

2.3 Scenario 3

While Scenario 2 defines the State Space based on point spread, Scenario 3 adds additional complexity by expanding the State Space to two dimensions: Point Spread and Number of Possessions. The strategy of going for two points can be highly dependent on the ammount of time remaining in the game. However, since time is a continuous space it was decided to discretize based on the number of offensive possessions, since it found that there are roughly 12 possessions per team during a football game (Sharp 2021).

The problem must further be represented as a two-player game wherein a home team taking an action (go for kick or two-point conversion) leads to the opponent then potentially scoring a touchdown on its drive and then attempting an extra point action as well. We therefore add 12 more drives such that odd-numbered drives represent the home team and even-numbered drives represent the opponent, taking actions back and forth until the 24th drive is completed. The State Space and Action Space are defined below:

$$\mathcal{S} = \begin{bmatrix} \texttt{drive} \\ \texttt{points spread} \end{bmatrix} \tag{3}$$

$$\mathcal{A} = \begin{bmatrix} \texttt{kick} \\ \texttt{two} \end{bmatrix} \tag{4}$$

The state transitions are constructed from a combination of the action space and the outcome of the opposite team's next drive. This structure is best demonstrated with an example–take that the game state is the first drive with a points spread of 6 (home team has just scored), i.e. State(1,6). The transition function will take on a categorical distribution over the four possible outcomes of the combination of action and next drive spaces. If the extra point is successful, 1 is added to the points spread. If the next drive is successful, we subtract 6 from the pointspread to represent the adversary's success. These are computed using Equation 1 for the extra point probabilities and p_d as the probability of a given team making a drive, fixed at 0.3 (a simulation assumption, see Section 5.2). Finally, each of the future states has it's `drive` value incremented by 1, representing the change in possession of the ball after an action is taken.

$$\begin{aligned} T(s, a, s') &= T(\text{State}(1, 6), \texttt{kick}, \begin{bmatrix} \text{State}(2, 6+1-6) \\ \text{State}(2, 6+1-0) \\ \text{State}(2, 6+0-6) \\ \text{State}(2, 6+0-0) \end{bmatrix}) \\ &= \begin{bmatrix} p \cdot p_d \\ p \cdot (1-p_d) \\ (1-p) \cdot p_d \\ (1-p) \cdot (1-p_d) \end{bmatrix} \end{aligned} \tag{5}$$

Similar distributions were used for the home team's `two` action and the opponent's `kick` action. The opponent is given a fixed policy of kicking by limiting the action space to $\mathcal{A} = [\texttt{kick}]$, therefore transitions from any state with an even drive value and a `two` action gave the terminal state a weight of 1, signaling that the action is not allowed.

The reward model for Scenario 3 follows from that of scenario 2. A notable addition is the reachability of win/loss states within the MDP. Given that the states represent progression in the game, as the MDP is solved, the win/loss state rewards are able to propagate to earlier drives in the State Space and thus incentives two-point conversions much earlier in the game.

3. Data Acquisition

Our various transition models rely on the probabilities of a specific NFL team's offense converting a two point conversion, a specific NFL team's defense stopping a two point conversion attempt, a specific NFL team's kicker converting a PAT kick and a specific NFL team scoring a touchdown on any given drive. While we initially investigated using an Web-based API from SportsRadar.com, the API proved to be an ill fitting for our needs. Thus we pivoted to accumulating the data we need into a csv file that could be accessed by our Julia programs when needed.

A specific team's probability of converting a PAT Kick was sourced from ProFootballReference.com (Reference 2021). The probability is inclusive of all extra point attempts from any kicker on the team in the 2020 NFL regular season. Moreover, a specific team's probability of scoring a touchdown during any given drive in the 2020 NFL season was sourced from SharpFootballAnalysis.com (Hribar 2021). Finally, the probability that a specific NFL team's offense converts a two point attempt in the 2020 NFL regular season as well as the probability that a specific NFL team's defense stops a two point attempt in the 2020 NFL season were both sourced from Teamrankings.com (TeamRankings 2021). In cases where a specific team did not attempt an two point conversion in 2020 or in cases where a specific team's defense did not face a 2 point attempt in 2020, data from 2019 was used.

4. Policy Formulation

4.1 Value Iteration

The first approach is to solve for the optimal policy exactly through Value Iteration, represented through the Bellman Equation, shown in Equation 6. With the Value Function $U_0(s)$ initialized to zero, the Bellman Equation will iterate until it reaches convergence (Kochenderfer, Wray, and Wheeler 2022). Given the size of the State Space, it was computationally determined that the value $U(s)$ converged around 100 iterations, but would be varied during simulation testing.

$$U_{k+1}(s) = max(R(s,a) + \gamma \sum_{s'} (T(s'|s,a)U_k(s'))) \quad (6)$$

Once $U(s)$ convergences, the optimal policy can be equation, shown in Equation 7:

$$\pi^*(s) = argmax_a(U^*(s)) \quad (7)$$

4.2 Q-Learning

The second approach is solving through Q-Learning, which is an approximation method that iterates and updates the State-Action Value Function Q after every iteration, without computing the Transition Function (Kochenderfer, Wray, and Wheeler 2022). Similar to Value Iteration, 100 iterations were computationally determined to reach convergence but would review Q-Learning at 50 iterations for additional comparison. The update for Q-Learning can be found below in Equation 9:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma max_a(Q(s',a')) - Q(s,a)) \quad (8)$$

In order to perform Q-Learning, some form of generative policy needs to be used. The chosen policy in this case was Epsilon Exploration Policy, which chooses a random action based on a hyperparameter, which in this case was set to 0.5 (Kochenderfer, Wray, and Wheeler 2022).

For Value Iteration and Q-Learning, both the discount factors were set to 0.9 (seek the later rewards of winning game), and the learning rate was set to 0.5 for Q-Learning. Scenario 1 will only need Value Iteration (small State Space, and deterministic, explained in 7.1), whereas Scenario 2 and Scenario 3 will employ both methods.

5. Simulation Formulation

5.1 Scenario 1

Due to the nature of the State Space, a game simulation cannot be performed because the MDP analyzes an instantaneous point during the game, not the duration of the game. Instead of running a game simulation, it was best to analyze by varying the p_{suc} for PAT kicks and two point conversions at a given p_{stop} value.

5.2 Scenario 2

With the State Space based only on point spread, the game simulation is constructed using two assumptions: 1) the game is set 12 possessions per team (as stated in 2.3) and 2) the probability that any team scores a touchdown is set to 0.3 (preset for simulation simplicity, based on success likelihood). A game starts with a points spread of 0 (initialization) and the home team in possession of the ball. Using the probability mass p and parameters p_{suc} and p_{stop} for each action, the simulation function randomly determined whether the team would score. The policy at the current state was then extracted; to transition to the next game state, it randomly determined if p was true or false (Bernoulli parameter); if true - based on the action taken (determined by the optimal policy), then the extra point would be achieved, if false, then it would result only as a touchdown (+6).

If the opposing team received the ball and scored, they would only result to PAT kicks. After running through 12 drives, the winner was determined on point spread, and would be redirected to the terminal state, ending the simulation. To simulate multiple games, the simulation function was iterated through k times (10,000 games). The win probability was determined by taking the cumulative average of wins to the total number of games at every 50 simulations.

5.3 Scenario 3

The Scenario 3 simulation follows the structure of Scenario 2. The game starts on drive 1 with a points spread of 0. Instead of 12 drives, there are 24: on odd drives, the home team action is extracted from the optimal policy; on even drives (opponent possession), the only action option is to kick. Determination of winner and simulation parameters also follows from Scenario 2.

6. Results

6.1 Scenario 1

By varying the p_{suc} value for each action at a given p_{stop}, the values can plotted for both PAT kicks and 2 pt conversions, as shown in Figures 1 and 2.

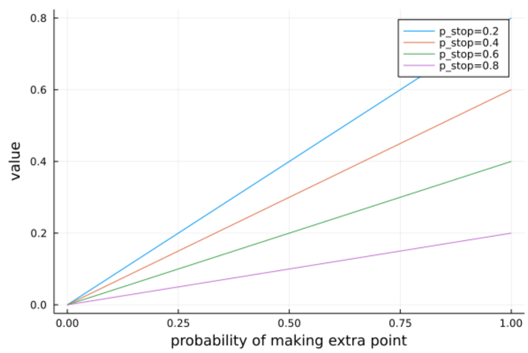

Figure 1: Value of PAT Kick in Scenario 1

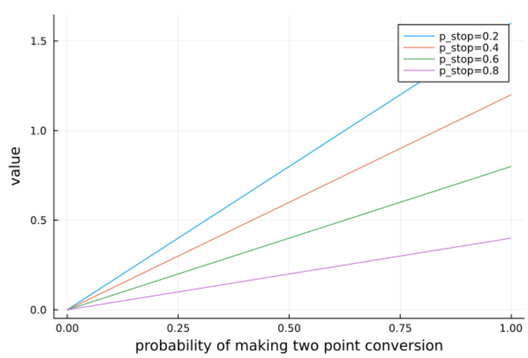

Figure 2: Value of two point conversion in Scenario 1

6.2 Scenario 2

With the simulations completed, the win probability was plotted for Value Iteration and Q-Learning Solving Methods, with Q-Learning varying between 50 and 100 iterations, as noted by the difference in Figure 3. With the simulations plotted, win probability for each Solving Method was calculated, with the results shown in Table 2.

6.3 Scenario 3

The Scenario 3 simulations again produced win probabilities over the Value Iteration and Q-Learning approaches. Figure

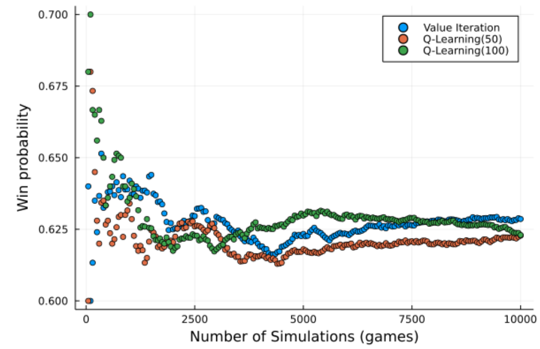

Figure 3: Scenario 2 Problem Setup with 10000 simulations using Value Iteration and Q-Learning at 50 and 100 iterations respectively

Algorithm	Number of Iterations	p(Win)
Value Iteration	100	.6274
Q-Learning	50	.6210
Q-Learning	100	.6289

Table 1: Win Probability for each Policy Solver using Scenario 2 Problem Setup

5 shows the win probability from simulation using a Value Iteration optimal policy with a varying max iterations parameter. Figure 6 compares the the Value Iteration with 100 iterations to the two Q-Learning trials. Table 2 displays relevant data on the algorithms.

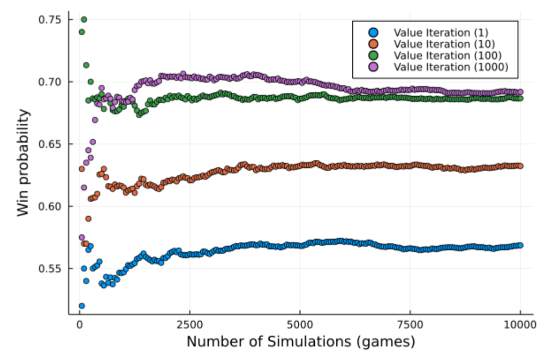

Figure 4: Scenario 3 Problem Setup with 10000 simulations using Value Iteration with Max Iteration parameter values of 1,10,100, and 1000

7. Discussion

7.1 Scenario 1

With a finite and discrete State Space, the input probability parameters will dictate the optimal action after scoring a touchdown. Looking at Figures 1 and 2, while also referring to statistical data of kicking, scoring and stopping percentages from the 2020 NFL season, NFL teams will generally obtain a value 0.7-0.9 (upper right quadrant of Figure

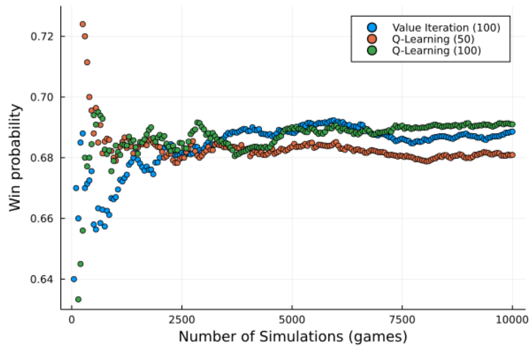

Figure 5: Scenario 3 Problem Setup with 10000 simulations using Value Iteration (100 iterations) and Q-Learning (50 and 100 iterations)

Algorithm	No. Iterations	$p(\text{Win})$	Runtime (s)
Value Iteration	1	0.561	20.86
Value Iteration	10	0.634	201.68
Value Iteration	100	0.687	1864.22
Value Iteration	1000	0.691	14855.72
Q-Learning	50	0.68	0.459
Q-Learning	100	0.68	0.715

Table 2: Relevant data for the algorithms implemented in Scenario 3's

1) because PAT kicks have been mostly successful, with the opponent typically unable to stop the kick. Compared with two point conversions in Figure 2, NFL teams will obtain a value less than 0.5 because they are much harder to make. Based on this information, the optimal action will always be to kick, unless the team has a far better two point conversion percentage.

While Scenario 1 attempts to answer the project objective, the finite and small discrete State Space limits the scope of the problem, and cannot determine whether the team will win or lose (looks at a single "snapshot" in the game). Thus, Scenarios 2 and 3 will be reviewed more closely since they account for game time factors (points spread and drive).

7.2 Scenario 2

While accounting for the complexity of the point spread during the football game, Figure 2 shows that both implementation of Q-Learning at various updates and Value Iteration return an average cumulative win probability of .62 vs the opponent's fixed policy of always kicking PAT kicks. While Q-Learning is an approximate Model-Free approach, it has fairly the same win probability, even at 50 iterations. This win probability is important because based on the criteria established by the Reward Model in Section 2.2, the team is going for more two point conversions than the opponent, especially when the game is at stake (down 1 point). By going for these close games, the optimal policies are resulting in more wins to losses vs a team who always kicks PAT's. Therefore, going for two points in tight point spreads (even when it's risky) is beneficial to winning the game, vs playing it safe in kicking extra points.

7.3 Scenario 3

Similarly to Scenario 2, both Value Iteration and Q-Learning given 100 iterations produced a policy with similar win probability (.68 and .687) as shown in Figure 5. Increasing the maximum number of iterations for Value Iteration to 1000 does slightly improve performance ($p(\text{Win}) = .691$), but at the price of a significant increase in runtime due to the algorithm complexity's linear dependence on k. Win probability performance showed asymptotic behavior with $k = 100$ demonstrating near-optimal performance with reasonable runtime as shown in Figure 4.

It is also of note that the winning percentages of the policies generated by Scenario 3 at 100 iterations were both higher than winning percentages of the policies generated at 100 iterations by Scenario 2 (.6274 and .6210 vs .68 and .687). This result makes sense as the State Space of Scenario 3 includes how many drives are remaining in the game where Scenario 2 does not include this information in it's State Space. The largest rewards are given at the end-game states–as the algorithms explore the state space, these rewards are able to propagate through Scenario 3 to influence policies in earlier drives differently than they do in later drives. This awareness of the remaining game horizon lends to accounting for a "sense of urgency" and therefore a higher degree of strategy, yielding improved performance and applicability relative to Scenario 2.

8. Conclusion

By examining the results of all three Scenarios, there are several conclusions that can be drawn from the results of this project. First from Scenario 1, it is clear that the value of a 2pt or 1pt attempt will always be somewhat tied to a team's ability to convert either type of attempt, regardless of game scenario. Common sense says there will situations in which this is overridden (For example, final drive of game when a team down by 8 points has just scored a touchdown to reduce the point spread to 2), however in general this tells us that if a team is bad at 2 pt attempts, they should just attempt 1 pt FG in most situations (and vice-a-versa). Since most kickers in the NFL have 1 pt extra point percentages in the 90's, while a team's 2pt extra is almost always lower, our results from Scenario 1 matches most NFL coaches' intuition as there are overwhelmingly more 1 pt extra point attempts in the NFL than 2 pt extra point attempts.

However, the winning percentages produced by following the policies generated in Scenarios 2 and 3 suggest a team's winning percentage could be improved if they followed a policy that was generated with Q-learning or Value Iteration. A winning percentage of .68 over the course of a 17 regular game season would result in 11 or 12 wins per year. A team that wins this many games is almost certain to qualify for the NFL post-season and have a shot at winning the Super Bowl.

9. Future Work and Expansions

If this work were to be continued, one possible improvement to explore is increasing the fidelity of Scenario 3 by replacing number of drives in the State Space with an actual time remaining in the football game. While using 12 drives per team was a basic way to simulate the transient nature of a football game, it does not capture any variation in a given team's offensive pace given the game situation or the average pace of their offense compared to the average of all other teams. For example, towards the end of the game, the winning team is often trying to waste as much time as possible during their own drive so that the team that is losing does not have enough time to score during their drives. This would require simulating(or randomizing based off of historical distributions) how much time each drive takes and then subtracting it from the time remaining in the game. It also introduces the issue of if there is less time remaining in the game then the simulated/randomized time of a drive. Does the drive automatically result in a reward of 0 or would the team potentially attempt a 3 point Field Goal given the point spread and where they were on the field? This would likely require adding additional states and fidelity to the game simulations, as drives resulting in Field Goals worth 3 points (rather than a 6 pt touchdown or 0 pts for no touchdown) were not explored at all in this paper.

The construction of the transition and reward functions also leave much to be tailored. The transition model relied heavily on simplifications of the game and the stats impacting the outcome of each action–as an example, we drew the success probability of a kick from the team's kicking percentage stats, but the health of the kicker, weather, etc. would also have an impact). The reward function used in this study is also only one of many possible approaches; others could be explored in future adaptations. Finally, the simulation assumed a constant expectation for drive outcomes (whereas a real game has different probabilities for each team) and an opponent with a fixed suboptimal policy (always kicking).

10. Contributions

Workload was distributed fairly evenly throughout the team. Tasking was generally distributed as follows: All members participated in the brainstorming and refinement of the idea. Michael Dacus designed and wrote the code for Scenario 1 and Scenario 2. Ian Hokaj designed and wrote the code for Scenario 3. Tyler Weiss performed the data collection and analysis. All members contributed to the report writeup.

11. Acknowledgments

The authors would like to thank Dr. Kochenderfer and the AA 228 (Decision Making Under Uncertainty) Course Assistant Team for providing feedback and insight for improving the Final Project.

12. Source Code

Source code, data, and figures can be found on the project GitHub repository: `https://github.com/mwdacus/AA-228-Final-Project.git`

References

Alder, J. 2020. Two-point conversion in football. available at: https://www.liveabout.com/about-football-glossary-two-point-conversion-1334140. accessed: 11.30.2021.

Brown, T. 2021. A brief history of football's extra point. available at: https://fieldsoffriendlystrife.com/2021/04/11/a-brief-history-of-footballs-extra-point/. accessed: 11.30.2021.

Hribar, R. 2021. What we can learn from team per drive scoring output for 2021. available at: https://www.pro-football-reference.com/. accessed: 11.30.2021.

Kochenderfer, M.; Wray, K.; and Wheeler, T. 2022. *Algorithms for Decision Making*. MIT Press.

Reference, P. F. 2021. Football stats and history. available at: https://www.pro-football-reference.com/. accessed: 11.30.2021.

Sharp, W. 2021. How many possessions can a team receive per game in the nfl? https://www.sportskeeda.com/nfl/how-many-possessions-team-receive-per-game-nfl. accessed: 11.30.2021.

Smith, M. D. 2021. Extra point success rate down to 92.9 percent, lowest since 1979. available at: https://profootballtalk.nbcsports.com/2021/11/22/extra-point-success-rate-down-to-92-9-percent-lowest-since-1979/. accessed: 11.30.2021.

TeamRankings. 2021. Sports, picks, predictions and data. available at: https://www.teamrankings.com/nfl/stat/two-point-conversion-pct?date=2021-02-08. accessed: 11.30.2021.

Tomlinson, S. 2015. Nfl's new extra-point rules a good 1st step, but there's still more to be done. available at: https://bleacherreport.com/articles/2470129-nfls-new-extra-point-rules-a-good-first-step-but-theres-still-more-to-be-done. accessed: 11.30.2021.

Urschel, J., and Zhuang, J. 2011. Are nfl coaches risk and loss averse? evidence from their use of kickoff strategies. *Journal of Quantitative Analysis in Sports* 7:14–14.