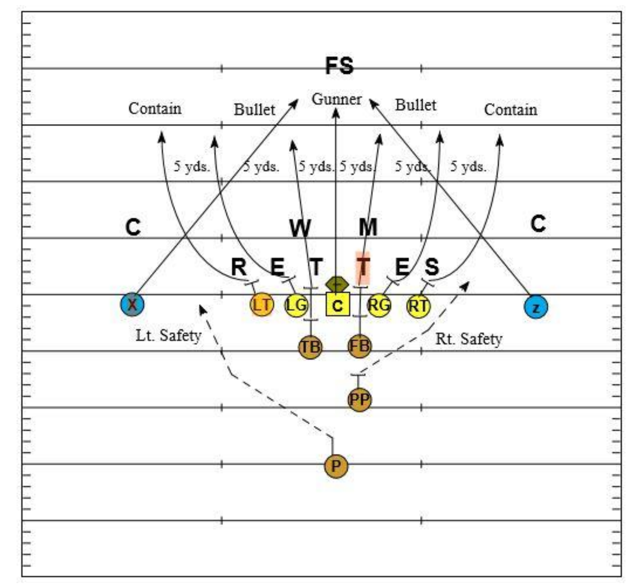
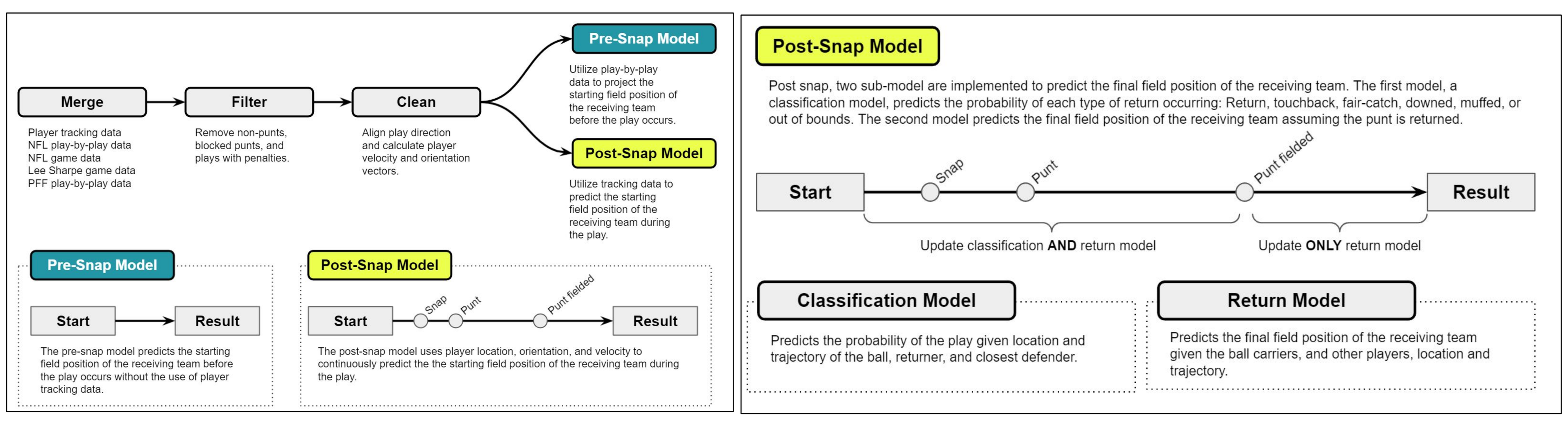
Expected punt outcomes in American Football

Digant Jani, Ryan Keeney, Salman Saeed, Shainu Prakash, Thanigaivel Shanmugam, Thang Trieu

Background and Motivation

In football, special teams are composed of members of the roster and they are placed in the game during kicking plays. Members of this group include kicker, punter, returner, and any members who are involved in the kick and return coverage. Our objective is to utilize time-continuous and incorporate spatial information - a first for special teams players to develop a novel approach that is suited for this play type and use it to evaluate the players and strategies that have not been studied previously at this level. We will further expand our evaluation to include players who are directly impacting the outcome of the play but may not have stats associated with their performance.
The motivation behind our objective is to bring to light the importance of special teams and by analyzing their plays, we hope to use the results to the team's advantage. For our analysis, we will focus on one type of special team plays, punts, and attempt to predict the starting field position of the receiving team's offense. We will leverage player tracking data from the NFL and player performance grading from Pro Football Focus (PFF) to develop a framework for analyzing punts including sub-models whose results are connected to an interactive Shiny web application running on an Amazon EC2 linux server. Our aim would be to predict the expected punt distance, return type and return distance at various points during a play. The success of the model would be in terms of predicting the expected end-of-play yard line.





Approaches

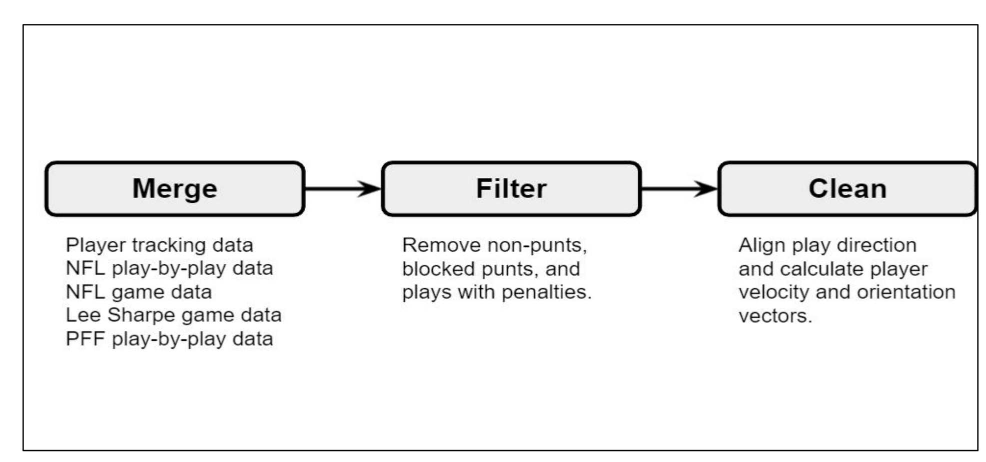
To establish a baseline, a Pre-Snap Model was developed which predicts the starting field position of the receiving team before the play occurs without the use of player tracking data. For this analysis, a random decision forest model was selected due to it's low bias. For punting, we would need to consider a classification model, and estimate the probability of a punt being returned, downed, fair-caught, kicked out of bounds or muffed, i.e. the punt receiver successfully catches the ball. The probability of each outcome could be estimated using a boosted random forest non-parametric model which will take inputs such as the punt receiver's speed and distances to the ball and the nearest defender.
Additionally, a second Post-Snap model will assume the punt returner has fielded the ball and predict the expected end-of-play yard line. This would be subject to the locations and trajectories of the punt receiver, teammates and opposing team members.We used a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework will also be evaluated via the XGBoost package in R. The results will then be collected and displayed in a Shiny R application running on a Linux Ubuntu EC2 server from AWS. The interface will allow users to dynamically explore the change in return probability and expected return distance dynamically throughout a play. Additionally, the results collected from the players performance-vs-expectation will provide a framework to evaluate and compare players.

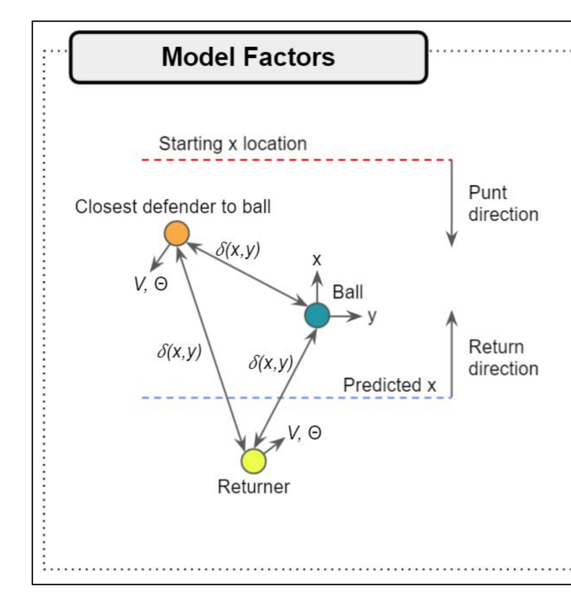
Data

Datasets were collected via downloads provided by the NFL, or were available publicly:

- Player Tracking Data: Ball and player location and trajectories collected via RFID at a frequency of 12.5 Hz during each play.
- NFL Play-By-Play Data: Play descriptions provided by the NFL, including the starting field position, result, and players involved in the play.
- NFL game Data: NFL game information, including team, location, and game outcome.
- Lee Sharpe game Data: Detailed NFL game information, including team, location, weather, game outcome, and betting market information.
- PFF Play-By-Play Data: Detailed play descriptions provided by PFF, including descriptions of personnel, scheme, and results for each play.

There are five-thousand valid plays in our dataset, resulting in over 13 million data points in the tracking data.





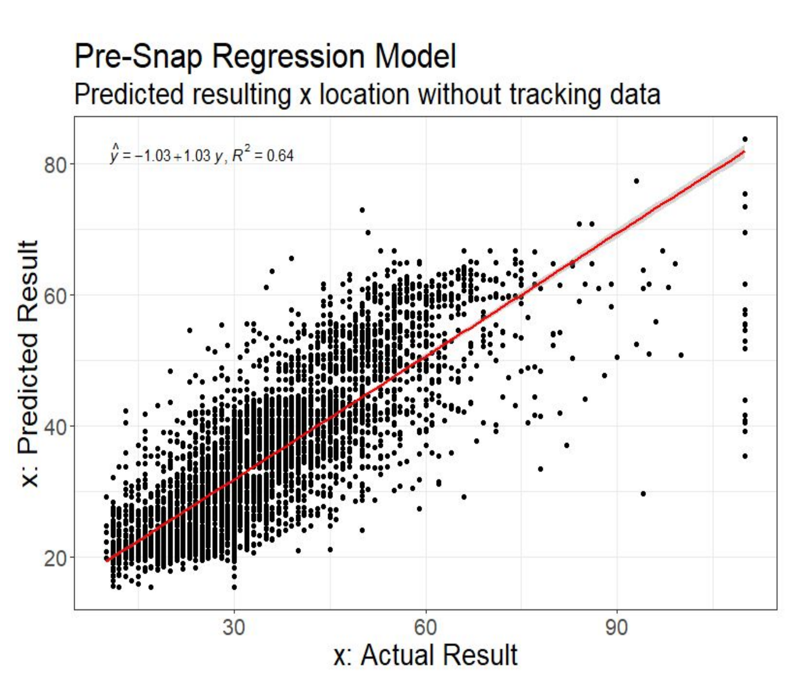


Figure 1

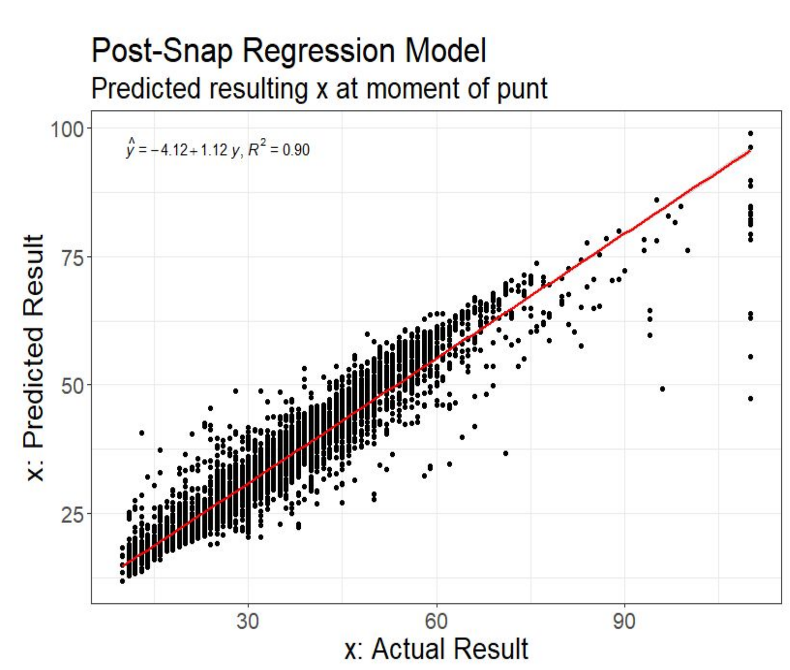


Figure 2

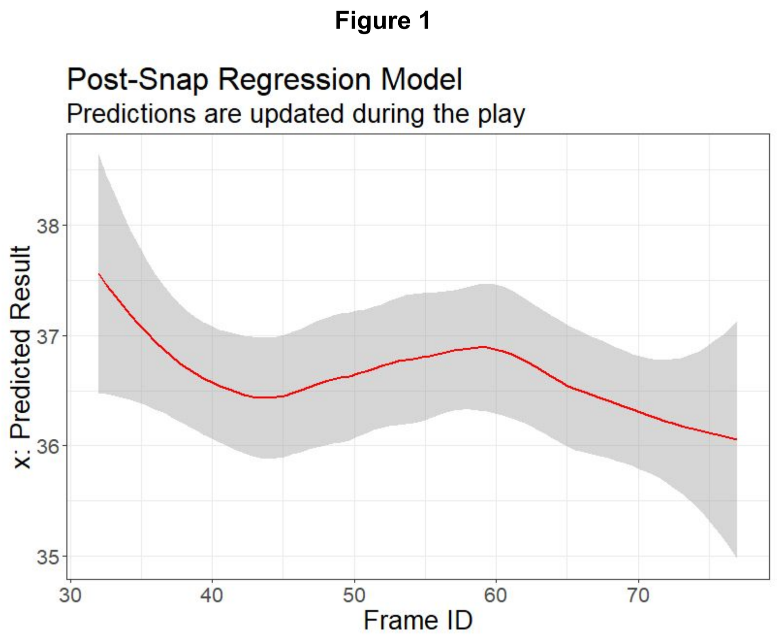


Figure 3

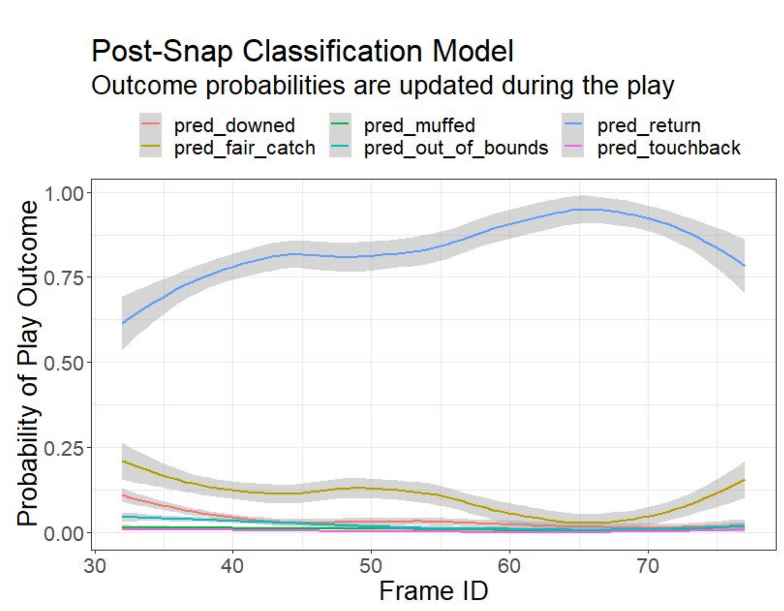


Figure 4

Experiments and Results

To establish a baseline, a Pre-Snap Model was developed which predicts the starting field position of the receiving team before the play occurs without the use of player tracking data. A simple random tree model using the factor set: distance to the opponent's end zone, number of gunners, number of jammer resulted in a RSQ 0.64, RSME of 9.97, and MAE of 6.88 on a 5-fold cross validation split 80/20 (training/testing) *[Figure 1]*.

We performed another approach called the Post-Snap model where we used player location, velocity and continuously predicted the starting field position of the receiving team during play *[Figure 2,3]*. Within the post-snap model there are two primary sub models. First, a classification model which predicts the probability of each type of return occurring: Return, touchback, fair-catch, downed, muffed or out of bounds *[Figure 4]*. Second, post-snap model predicts the final field position of the receiving team assuming the punt is returned. The classification model gives the probability of play given location, trajectory of the ball, return and closest defender. Return model predicts using the ball carriers, and other players, location and trajectory. This will evaluate multiple independent and dependent model aspects of the model and predict the final output of punts using that output. Based on Analysis we performed a few experiments and evaluated models based on it to find out what is needed for optimization and tuning the approach.

Our model predicted the punt outcome classification (Post-Snap model) with an accuracy of 91.1% for plays from the testing dataset.

Web App Link

Shiny App URL:
https://bit.ly/GT_Team123/

Application Demo

This lightweight demo has 4 stored plays to explore (A single play can have up to five-thousand data points). Select a game and play id and an animated plot will show the the play as described by the player tracking data. Two predictions are show: (1) A pre-snap prediction created using non-player tracking data in RED, and (2) A post-snap model which predicts the final location between the time of the punt and the next event in BLUE. See result in *[Figure 6]*.

This approach allow for additional insight into player performance and strategy. For example, the average yards over expectation can be calculated based on the predicted starting field position at the time of the punt. In this table, we explore which returners advanced the ball further than the prediction. This is an example output, and while a positive average over expectation may be attributed to good decision making or skill, this analysis is not intended to be a comprehensive analysis of player performance - it does demonstrate the ability to be building block for more advanced and accurate player and strategy evaluation. (Minimum 50 punt returns, Seasons 2017-2020). *[Figure 5]*

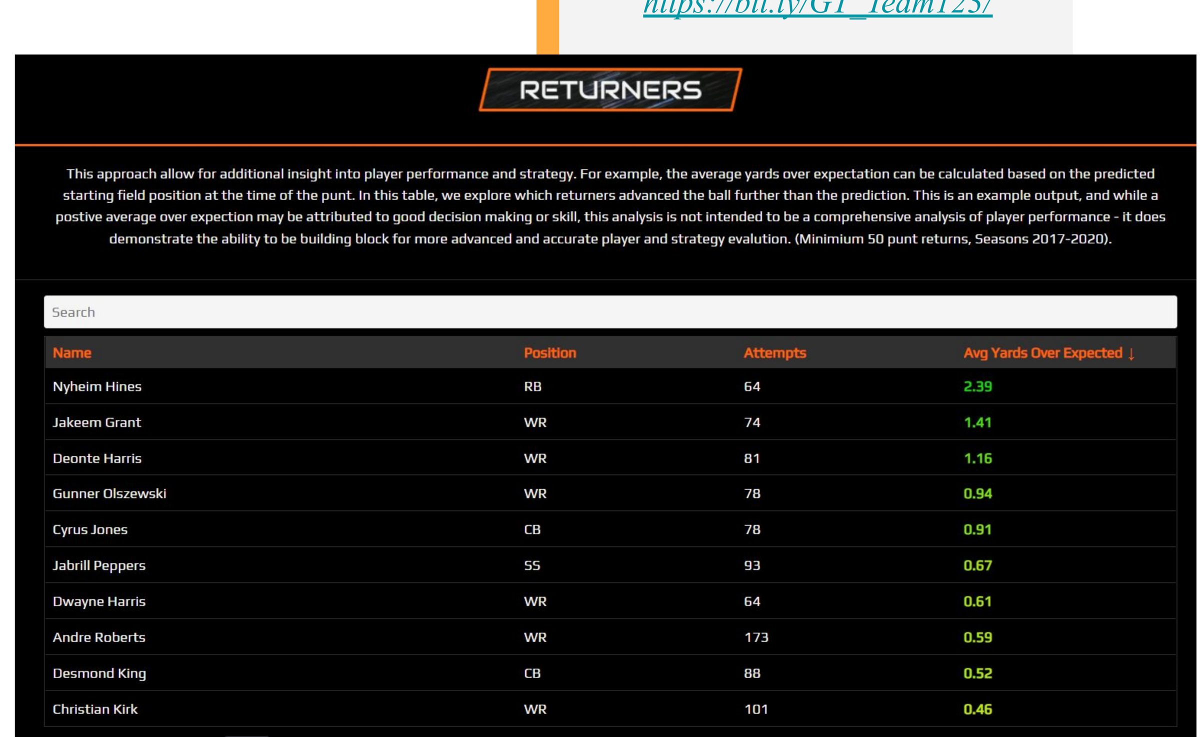


Figure 5

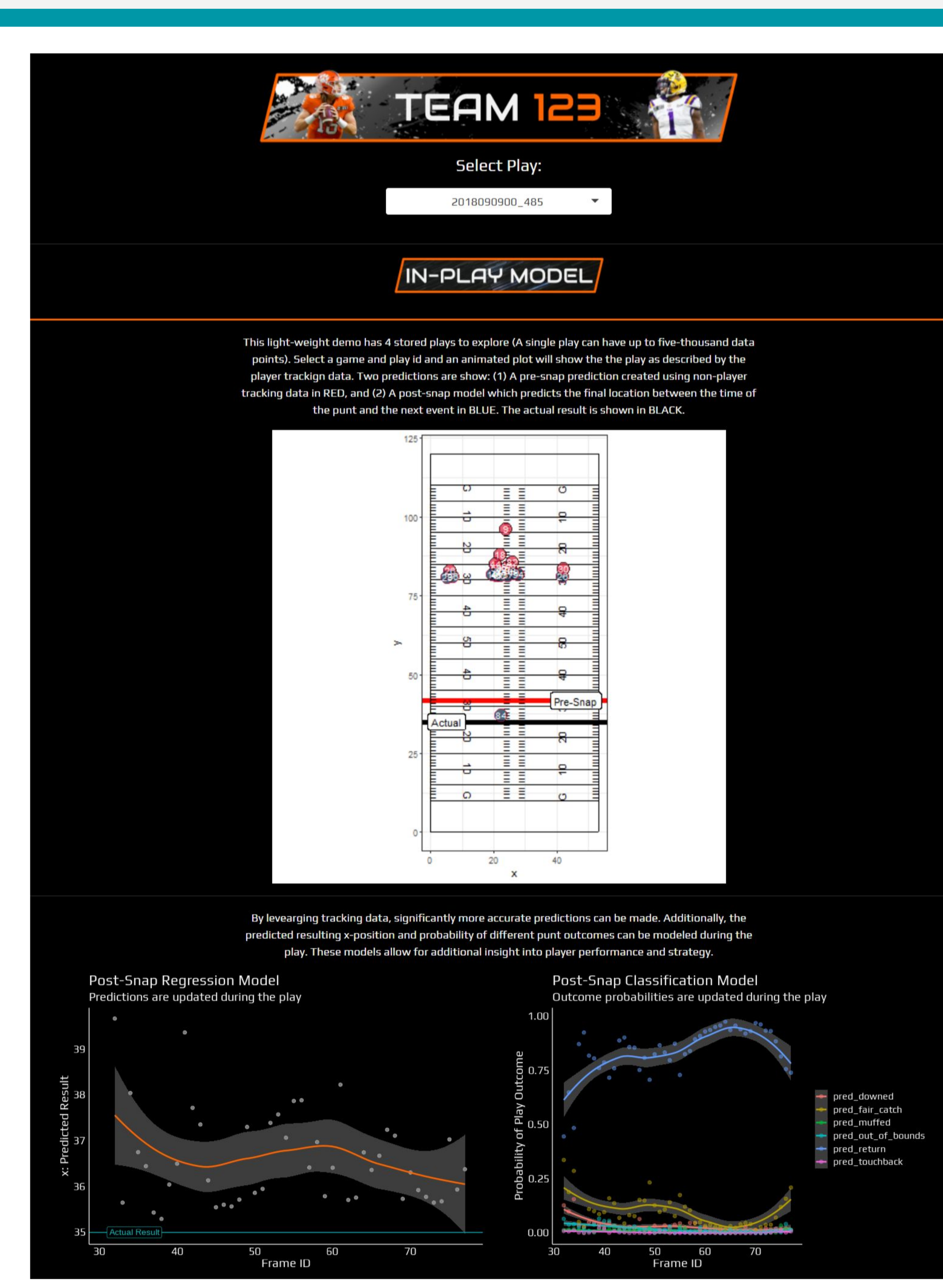
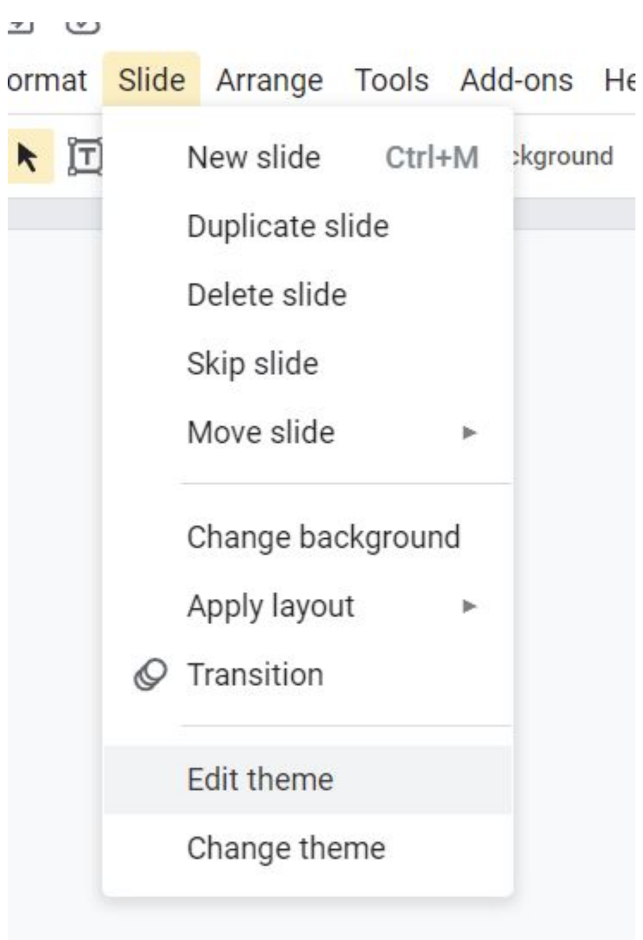


Figure 6

Reference



- Size is already set to 30x40
- Min size font: 18
- If you want to edit **background** colors/shapes, try editing the master slides

Examples:

- example poster 1
- example poster 2
- example poster 3