

Proposal for Springboard Capstone 2

Chris Lomeli
January 2023 cohort

Problem Statement

During an NFL football game the offensive coaching staff would like to predict the next best play to call for each situation that comes up

Context

American football is played across the US in parks and schoolyards as a physical sport. But as the game progressed through colleges and professional leagues it became highly strategic.

We are creating this model for an imaginary problem. In current reality, the offensive coaching staff would not use an ML model. In an effort to preserve the game as an athletic event, and avoid turning it into a technology competition, the NFL has closely managed the introduction of machine learning.

Criteria for success

Given a set of conditions (the situation) - the model will predict the 'best' next play to call, based on historical data and currently available statistics.

Scope of solution space

The model will be focused on predicting offensive play. That's really only half the game - defensive strategy plays a huge role in determining wins and losses. I can't think of a way to handle defensive strategy in the timeframe we have.

There are many factors that could play in the final prediction - such as the health and stats of individual players. We will not go into great depth there - the supposition being that a team can under-perform even if it has great players - so we'll focus a bit more on the team performance. This first approach may or may not play out the way we expect.

Constraints

We are limited by data - professional data is expensive, so we are using free data
We are limited by time - the timeframe is 2 months or less

Proving that our prediction is correct is difficult because in reality we can't run the next play several different ways to verify that our prediction was correct and the other options are incorrect. We may be able to compare the model's output against existing probability models - but that is a far cry from proving that the model is 'right' in a real situation.

From my limited understanding, this constraint appears no different than other ML examples, including the guided capstone we've just completed

Stakeholders

Springboard capstone
NFL offensive coaching staff

Data sources

Next gen stats	GitHub - cooperdfff/nfl_data_py: Python code for working with NFL play by play data.
Spreadspoke score data	NFL scores and betting data Kaggle
FiveThirtyEight data	nfl-elo
Kaggle NFL Play by play data This static set looks like a version of the next gen stats data (below) but is a bit more messy - good for cleaning.	Detailed NFL Play-by-Play Data 2009-2018 Kaggle

Feasibility of our approach

It's arguably feasible to create an ML model for playcalling. The approach will be something close to the following:

Dependent variable:

- is the best play to run next. This is our output.

Potential independent variables:

- What Quarter are we in?
- What Down are we on?
- Seconds to the end of the quarter
- Seconds to the end of the game
- The current score
- Yards to go to get to a first down or a touchdown

- Opposing team in our division or not
- Our record against the opposing team in this season

Rollups

Data that is summarized from previous records

- Our offensive and defensive performance this season
- The opposing team's offensive and defensive performance this season
- Our team stats this season
- Our Passing, Rushing, Field-goal, seasonal performance
- Bench stats - who have we lost? Who is playing injured
- Passing, Rushing, Field-goal, seasonal performance

References

Existing projects

Code example	NFL Betting Model Kaggle
Can you beat 538 predictions?	ukritw/nflprediction: Predicting NFL games using Machine Learning
Article - Remedial, but uses the play by play dataset	How to Win in the NFL with Machine Learning by Rich Folsom Towards Data Science

Existing research

2019 MIT Thesis	Leveraging Machine Learning to Predict Playcalling Tendencies in the NFL Udgam Goyal
2020 Stanford Thesis	Deep Learning for In-Game NFL Predictions
Berkley -School of Information Talk	Using Machine Learning for Predicting NFL Games Data Dialogs 2016

	Predict Football Match Winners With Machine Learning And Python
Article - Remedial, but uses the play by play dataset	How to Win in the NFL with Machine Learning by Rich Folsom Towards Data Science
Git example play by play example in R	nflscrapR/scrape_play_by_play.R at master
Using Machine Learning for Predicting NFL Games	Using Machine Learning for Predicting NFL Games Data Dialogs 2016
Predicting Point Spread in NFL Games:	Predicting Point Spread in NFL Games CS229
Getting next gen stats	https://www.youtube.com/watch?v=wWgGgmqijNU