

NFL Play Predictor

Viraj P. Modak

Problem statement:

In any sport a team's performance can be said to primarily depend on 3 aspects: Strategy, Tactics and Execution. We can explain this for National Football League (NFL) as follows. Prior to the game, the team decides on an overall strategy for it. During the game, based on the situation the offense then makes tactical decisions and calls plays. The players then try to execute them on the field. And the objective of the defense is to disrupt that execution or at least minimize its impact. In such a scenario, if the defense already knows what play is going to be called, that can be a huge advantage.

There are multiple factors which can decide what offensive play is going to be called, for example, the play clock, down, yards to go, yards to goal etc. Except in obvious situations, it is not always straightforward for humans to process all this information and predict the play-call. And the objective of this project is to solve this problem using machine learning (ML).

Target client:

The target client for this project will be NFL coaches, in particular the defensive coordinators and their support staff. In the absence of a predictive model, coaches will have to rely on human judgement to anticipate the play-call. A human brain equipped with game knowledge and experience may well be able to do that. However there are often situations with a degree of uncertainty. In such cases data-driven decision making is the best approach that one can take, where the results from this project can come in handy. An added advantage of an ML algorithm is that it can objectively quantify weights of the individual factors which influence the final play, whereas a human brain cannot.

Data to be used:

For this purpose, the "**nflscrap-R play-by-play**" data will be used which is available freely on GitHub. Data is recorded for every NFL game (regular and post-season) from 2009 through current. Analysis will be restricted to the end of 2018-2019 season. Data can be found here;

https://github.com/ryurko/nflscrapR-data/tree/master/games_data/regular_season

Problem solving approach:

The final goal of this project is to have a predictive model which can predict a Run/Pass play based on the game situation. The game situation can be summarized using multiple variables including*:

1. Down
2. Yards to go
3. Field position (Yards to goal)
4. Play clock
5. Score line
6. Play formation
7. Team pass performance in the game
8. Team run performance in the game
9. Dual-threat rating of the quarterback (QB)
10. Timeouts remaining

**the final set of variables eventually used may not exactly reflect the list here*

At a high-level, the following approach will be taken to solve this problem. Data will be cleaned up to a format which lists out the individual variables and the final outcome. Some of the factors e.g. QB dual threat rating will have to be calculated independently. Exploratory data analysis will then be performed to identify trends in those variables. The data will then be used to train and optimize a model using one or more ML techniques. Finally performance of these techniques will be evaluated.

Although the final objective of this project is to design the most accurate model possible, because of the unpredictable nature of sports there will be a ceiling to the performance level. Efforts will also be made to understand the gap - for example is there a trend in performance vs. the offensive team? In other words, are some teams more predictable than others? Does the performance change when the QB has a high dual-threat rating? And so on.

Deliverables:

The deliverables will include all the relevant code (Python) files and a project report document which summarizes all the key findings. Supplementary information and statistical data will be included as an appendix. The report is currently proposed to be in the form of a document but eventually can be changed to a slide deck.