Does the NFL have sufficient data to predict a first down with a proposed running play?

# Introduction

In 2015, the National Football League (NFL) placed computer chips in all its players in order to track all aspects, including position, movement, orientation and speed. The idea of it was, while working with Microsoft, to create an app for the Microsoft X-Box One video game console for users to see instant replays with instant data for people that don’t have access to watch the game. In 2018, The NFL reached a deal with Amazon to not only test out streaming Thursday night games on Amazon’s game streaming platform Twitch, but to utilize the data collected from the micro chips to implement big data analytics. In 2019, the NFL placed two data sets on Kaggle as a way to extend the data to aspiring data scientists. One of the data sets was focused on location of players and information about injuries, in a way to identify the number of factors that could be contributing to the number of players injured or concussed on the field. The other one, which will be the focus of this report, is a data set of all running plays made from Week 1 of the 2017 season to week 12 of 2019. This study will be looking at the factors given, and regressively looking for how many provided factors contribute to a successful play that leads to a first down. This study will also look at creating a predictive model to see if we can use the same fields and attempt to accurately predict if a first down will be successful.

# Literature Review

Write summary of the related papers that you reviewed here. Write the summary in your own words—don’t use the technical jargon from the paper that you don’t understand. Keep this section short—a short paragraph or few sentences about each paper you reviewed should be sufficient.

# Dataset

The data set was provided by the National Football league and posted onto Kaggle on November 2019, for a contest that lasted a month. 15 winners were selected for their use of Machine Learning and visualization of the data. In the data set provided there were 45 attributes provided and 682154 records provided. These records represent the 688 games played, and each game contains a total of 31007 plays, and every play had 22 players on the field, 11 for the defense, and 11 for the offence, including the player running the ball.

# Approach

Create a block diagram for the steps of your approach to clearly provide an overview. For example, if you first scrapped twitter, second applied NLP techniques to extract keywords, third labelled the tweets as positive and negative using a set of keywords, and fourth build a classifier, then you should create a box for each of the steps with arrows connecting one step to the next one. A sample block diagram is shown below.

Once this is done, explain each of the steps in detail. What are you planning to do in each step or have already done. For example, in the above case you would create subheadings for each of the steps.

## Step 1: <Name of the step>

Write details of the step 1. If there is any source code that you’d like to share then provide the link of the Github.

## Step 2: <Name of the step>

Write details of the step 2. If there is any source code that you’d like to share then provide the link of the Github.

………………………………………………………………………………………………………………………..

………………………………………………………………………………………………………………………..

## Step N: <Name of the step>

Write details of the step N. If there is any source code that you’d like to share then provide the link of the Github.

# References

Moynihan, T (2015, August 7) *All NFL Players Are Getting RFID Chips This Season* Wired <https://www.wired.com/2015/08/nfl-players-getting-rfid-chips-season/>

Dastin J, Paul A (2018, April 26) *Amazon, NFL reach $130 million streaming deal for Thursday night games: source* Reuters https://www.reuters.com/article/us-nfl-amazon-com/amazon-nfl-reach-130-million-streaming-deal-for-thursday-night-games-source-idUSKBN1HX3EP