**Predicting Run or Pass probability in NFL game**

* **Problem Definition**: This project is an analysis of the two most predominant plays in any given football game. The forward pass and the run also called a rush are the most common offensive plays during any given game. This analysis is endeavoring to determine which of these play types has the greatest probability of success and which will be executed next during a game. Having this information would be a sizeable advantage to both the offensive defensive units on the field. If NFL defenses are able to predict the next play call they will game plan to stop the offensive play call. On the other hand offenses will call the play with the greatest probability of success. Using past results from previous NFL seasons, predictions will be made to determine the probability of success and of the next play type. Independent Variables include the Down, Yards to go, Yard line, Play direction and Yards gained. The dependent variable is the Play Type either a run or a pass.
* **Background/literature review:** The NFL is the most popular league in the United States.It is a billion dollar industry with TV contracts with the largest TV distributers in the United States. Having a model to predict the probability of run or pass will enable teams to be more competitive through preparation. Moreover, individuals who gamble on NFL games will have a better opportunity at accurately predicting the outcome of games. Thus, having a predictive model to determine the play type is a great idea. This will give those who wager on the matches an outlook as to the predictions of the upcoming games and which team to place their wager on.
* **Data mining technique:** Logistic regression and Decision Tree Classification

**Machining learning technique:** Artificial Neural Network (ANN)

* **Description of design:** For predicting probability of run or pass, Logistic regression, Decision Tree Classification and Artificial Neural Network were used. Based on which model gives the best predictions, that model will be implemented. For predicting probability of run or pass, the most optimal model to use is Decision Tree Classification. Based on Logistic Regression, the paly type has been predicted with 64.6% accuracy.
* **Explanation of data (file format, independent/dependent variables, source)**

**File format:** The data is structured in a CSV file format, so it can be imported into the model.

**IDV:**

* **quarter**

1. Description: It represents the time played by quarter
2. Data type: Numeric

* **month**

1. Description: It represents the months from September - January
2. Data type: Numeric

* **ydstogo**

1. Description: It represents the amount of yards to gain for a first down
2. Data type: Numeric

* **Week**

1. Description: It represents week game is played (NFL season has 17 weeks)

2. Data type: Numeric

* **Season\_year**

1. Description: It represents year in which game is played

2. Data type: Numeric

* **drive\_id**

1. Description: It represents the drive (possession) number within each game

2. Data type: Numeric

* **Time**

1. Description: it represents amount of time remaining in a given game

2. Data type: Numeric

* **Pos\_Team**

1. Description: It represents the team in possession of the ball

2. Data type: Textual

* **Pos\_division**

1. Description: It represents division of offensive team

2. Data type: Textual

* **Def\_Team**

1. Description: It represents the defensive team

2. Data type: Textual

* **Def\_division**

1. Description: It represents division of defensive team

2. Data type: Textual

* **YardLine**

1. Description: It represents Yard line on the 100 yard field where the ball is located

2. Data type: Numeric

* **Down**

1. Description: 1-4 opportunities for team in possession of the ball

2. Data type: Numeric

* **Play\_direction**

1. Description: It represents the direction in which play is made on the field

2. Data type: Textual

* **yards**

1. Description: It represents yards gained or lost on any given play

2. Data type: Numeric

* **Game\_time**

1. Description: It represents if the game is played on regular time or prime time

2. Data type: Textual

* **home**

1. Description: It represents home team

2. Data type: Numeric

**DV**: **Play Type**

1. Description: It represents Run or Pass
2. Data type: Textual

**Source:** <https://www.kaggle.com/maxhorowitz/nflplaybyplay2009to2016>

* **Identification of Training and Testing data**

**Training set:** Data set is split in to 80% to create the training set i.e. (171,108 rows of 19 variables).

**Testing set:** Data set is split in to 20% to create the testing set i.e. (42,777 rows of 19 variables)

* **Testing and Validation Procedure:** After training the model with the training test. The model will predict the test set and the confusion matrix will be created to compare with the actual data. The model should predict play type probability with relative certainty.
* **Results and incites:** The results from the test set had compared with the actual scores and the result from the data set**.** The DTC model was able to predict the outcome with 75% accuracy.
* **Conclusion:** After performing the analysis on probability of success and predicting the play type our assumption did not meet our expectation. We needed to figure out why the run had a significantly higher probability of success over the pass. After looking at our data once again we were able to determine that the run had a higher probability because of the incomplete passes of the Quarter Back. The best QB percentage in the NFL averaged about 70% during the years 2009-2015. The percentage of successful handoffs from QB to running back was upwards of 95%. Thus the result of the run having the higher probability of success.
* **Citations**
* Pablo Bosch.(February 2018). Predicting the winner of NFL-games using Machine and Deep Learning.

Retrieved from https://beta.vu.nl/nl/Images/werkstuk-bosch\_tcm235-888637.pdf

* Zachary Owen and Virgile Galle.(2017). Predicting the NFL.

Retrieved from <https://docplayer.net/56085578-Predicting-the-nfl-zachary-owen-and-> virgile-galle.html

* Austin Grosel. Predicting NFL Coaching and Team Based Play Calls
* Orr, Connor. “NFL's Most, Least Balanced Offensive Play-Callers.” NFL.com ,

Retrieved from www.nfl.com/news/story/0ap3000000816875/article/nfls-most-least-balanced-offensive-playcallers .

* Jones, Willie D.(August 2015). Model Predicts Whether NFL Teams Will Run or Pass.

IEEE Spectrum: Technology, Engineering, and Science News.

Retrieved from spectrum.ieee.org/tech-talk/geek-life/tools-toys/statistical-model-predicts-whether-nfl-teams-will- run-or-pass.