Using Machine Learning and Non Game Factors to Predict the Outcome of NFL Games

Jacob DeMuth   
*Purdue Northwest Technology Department*  
Purdue NorthwestHammond, Indiana  
demuthj@pnw.edu

# Introduction

## Background

# The early start of using statics to predict NFL games was in 1980 when Harville used a linear method to try to make predictions. Their work was based on building blocks from their previous research in 1977 geared toward high school and college level football. AI techniques have become popular since the start of the 2000s and these models include Bayesian models and decision trees. While the original baseline was set by Boulier & Stekler at 61% and bookmakers at 65.8%, more recent studies done by the University of Southampton show that using Naïve Bayes an accuracy of 67.53% is the new baseline.

## Objectives

The objective of this research is to determine the impact of weather and stadium data on the outcome of which team wins in an NFL game.

## Significance of Project

The significance of this project is that a new area of potential features will be explored. Past literature has focused on game features such as offensive yards per game or number of defensive stops on average. By including features such as stadium elevation and temperature the question will finally be answered as to whether external factors have a significant impact on sports games.

# Literature Review

## Previous Works

One of the first mathematical methods introduced for predicting NFL games was proposed by Bill James in 1981. This method was called the Log 5 method. While this was primarily used in baseball the theory behind it could still be used in any sports game.

The first research paper involving the use of machine learning to predict the outcome of sports games was published by Purucker in 1996 [4]. This study involved the use of a machine learning technique known as neural networks. The neural network used predicted the outcome of 28 matches with a correct prediction for 21 of those games. While this was an important first step in the field, there were still plenty of improvements to be made. Purucker himself notes that the amount of data and features used is small meaning that there was still plenty of room for improvement.

In 2003 Kahn improved upon Puruckers neural network by including more features in his network and predicting more NFL games. A key difference with Kahn’s neural network was the use of 3-week historical averages as well as the overall season average. This allowed his model to achieve a higher overall accuracy. He also tested his model on 208 matches compared to Truckers 28 matches. Kahn was able to achieve an accuracy of 75 percent. In Kahn’s study he used weeks 1 through 13 of the 2003 season were used for training and weeks 14 and 15 were used for testing. He noted in his work that the inclusion of Vegas odds for matches and expanding the training data could allow for a better overall model.

Then in 2011 Jim Warner applied a machine learning technique known as the Gaussian model to predict NFL games [5]. His accuracy over the 2011 season was around 64.36 percent. He trained his model on data between the years 2000 to 2009. Another machine learning technique used around this time was called the Support Vector Machine model. This technique implemented by Albert Shau achieved an accuracy of 68.4 percent for the 2011 season. His model trained on data from 1970 to 2011.

Eventually, a paper published in 2018 by Pablo Bosch would go on to use deep learning to predict the outcome of NFL games [2]. The data used in Pablo’s research was from the years of 2009 to 2016 and the models implemented included SVM, Regression, Random Forest, Neural Networks, and LSTM. The accuracy achieved with these models averaged around 63 percent with Logistic Regression having the highest accuracy at 63.33 percent. Pablo noted in his paper that he felt improvements could be made by including data prior to 2009 and implementing features related to the players themselves such as how many pro bowls a player has or how many mvps they have.

In 2020 a paper written by Ryan Beal, Timothy Norman, and Sarvapali Ramchurn implemented more machine learning models than previous papers [1]. The new models implemented by them included Nearest Neighbors, Decision Tree, Random Forest, AdaBoost, and QDA. Their findings indicated that Naïve Bayes was still the best model achieving an accuracy of 67.53 percent over the 2015 to 2019 seasons. They noted in their research that while this accuracy is higher than bookmakers in Vegas, the model’s accuracy was not consistently higher than the bookmakers every year. Bookmakers gather their odds by using statistical and market demand methods. They conclude in their paper that the best performing model year over year would switch between Naïve Bayes and AdaBoost.

The key findings from past literature are that the field of sports prediction is still growing with every passing year. Previous papers have explored methods and techniques including neural networks and Naïve Bayes. Features such as historical averages or ELO ratings have been explored to try to achieve better overall accuracy in predicting the outcome of games. My research plans to further the field by exploring features and data not included in previous works. I plan to implement features including the weather, stadium surface, and other factors that have not been previously included in past works. The overall goal will be to achieve a higher overall accuracy with this new idea.

# Methodoloy

## Description of Methods

The methods for testing the impact of non-game related features on the outcome of NFL games was achieved by implementing these features in several machine learning models as well as calculating the correlation coefficients of these factors on the outcome of games. The coefficient calculation utilizes a -1 to 1 scale with the closer to -1 or 1 being a strong correlation. The results of the model will be compared to the results of using the ELO method for determining a likely winner of an NFL game. The metrics used will be ROC AUC, Brier, and Accuracy. The accuracy of the model would be determined using probability. The model will output the probability that it feels of the home team winning a game. If the probability is greater than 50 in favor of the home team and the home team wins, this is counted as correct. When the home team loses, and the probability is greater than 50, this is counted as wrong. The same methodology is applied when the probability is less than 50. When the probability is less than 50 and the home team loses, this is counted as a correct answer. When the probability is less than 50 and the home team wins this is counted as a wrong answer.

## Data Collection Process

Data was collected by using previous datasets hosted by FiveThirtyEight on GitHub which included game data and ELO data. Weather and missing game was collected through web scraping Pro Football Reference and the website NFL Weather Data. Historical Data between the years 2002-2022 was collected with partial 2023 data appended to the dataset. Training data used the 2005-2015 NFL season with testing including the 2016-2023 data. This data was stored in CSV format.

## Tools Used

The programming language that was used for this was python version 3.7.4. The libraries utilized were as follows: Pandas, NumPy, Datetime, XGBoost, OS, Lazy Predict, Matplotlib, Sklearn, Lightbgm, and MLX tend. This was all complied together with a program called Jupyter Notebook that functioned as the IDE. The models tested were as follows: Random Forest Classifier, Extra Trees Classifier, LGBM Classifier, XGB Classifier, Baggin Classifier, Logistic Regression, KNeighbors Classifier, Gaussian Naïve Bayes, Decision Tree Classifier, Quadratic Discriminant Analysis, SVC, Ridge Classifier, and SGD Classifier. Excel was used to store the data.

## Features Used

The features that were focused on for this research were stadium capacity and elevation as well as the temperature and wind speed of the day a game occurred. The following are all the features utilized in this project along with a brief description of what they are:

spread\_favorite

This is the spread that Vegas had set for a game.

over\_under\_line

This is the over under line set by Vegas for a game.

weather\_temperature

This was the temperature, in Fahrenheit, of a game that occurred.

weather\_wind\_mph

This was the wind speed, in Miles per Hour, of a game that occurred.

hm\_avg\_pts\_diff

Difference in points scored by home team minus points scored by the away team.

aw\_avg\_pts\_diff

Difference in points scored by away team minus points scored by the home team.

hm\_avg\_pts

The average points a team scored at home.

aw\_avg\_pts

The average points a team scored when not home.

.

stadium\_neutral

Notes if a particular stadium is the away stadium for both teams.

home\_favorite

Notes if the home team is the favorite team as determined by Vegas.

away\_favorite

Notes if the away team is the favorite team as determined by Vegas.

elo1\_pre

The ELO rating of team 1 before a game.

elo2\_pre

The ELO rating of team 2 before a game.

elo\_prob1

The probability, based on team ELO, of team 1 winning.

elo\_prob2

The probability, based on team ELO, of team 2 winning.

result

Notes if the home team won or lost the game.

hm\_avg\_pass\_yds

The average amount of offensive passing yards when the team is at home.

aw\_avg\_pass\_yds

The average amount of offensive passing yards when the team is away.

hm\_avg\_rush\_yds

The average amount of offensive rushing yards when a team is at home.

aw\_avg\_rush\_yds

The average amount of offensive rushing yards when the team is away.

hm\_avg\_total\_yds

The average amount of total offensive yards when the team is the home team.

aw\_avg\_total\_yds

The average amount of total offensive yards when the team is the away team.

hm\_avg\_rushing\_attempts

The average number of rushing attempts of the team when they are home.

aw\_avg\_rushing\_attempts

The average number of rushing attempts the team has when away.

hm\_avg\_fumbles

The average number of fumbles when the team is at home.

aw\_avg\_fumbles

The average number of fumbles the team has when away.

hm\_avg\_int

The average number of interceptions thrown by the team when they are the home team.

aw\_avg\_int

The average number of interceptions thrown by the team when they are the away team.

hm\_avg\_turnovers

The average number of turnovers of the team at home.

aw\_avg\_turnovers

The average number of turnovers when the team when not home.

hm\_avg\_drives

The average drives down the field of the team at home.

aw\_avg\_drives

The average drives down field of the team when not home.

,

stadium\_capacity

The seating capacity of the home stadium.

stadium\_elevation\_meters

The elevation, in meters, of the home stadium.

# Data Analysis

## Findings

When compared with using a standard ELO rating the models’ overall ROC AUC score was 0.7028 compared with ELO’s 0.6871. The Brier score for the models was 0.2184 with ELO’s Brier score being .2243. Since a higher ROC AUC score and lower Brier score are better, the models have outperformed ELO.

The accuracy of the models was 0.6455 compared to the ELO’s accuracy of 0.6283. While the models’ accuracy outperformed ELO’s, the accuracy overall did not outperform previous works.

When comparing the impact of the stadium and weather features on the outcome of NFL games there appears to be no correlation. Stadium elevation had a correlation of 0.02, capacity had a correlation of 0.01, wind speed had a correlation of 0.02, and temperature had a correlation of -0.04.

## Data Visualization

Scores

A screenshot of a computer

Description automatically generated

Correlation Coefficient Heatmap

A green and yellow squares with black text

Description automatically generated

## Interpretation of Results

These results indicate that game features have more of a impact than non game features. Despite the addition of weather and stadium-based features the overall accuracy of the model was worse than that of previous works.

# Discussion

## Analysis and Implication of Findings

These findings indicate that the efforts of researchers are better spent on game related factors rather than factors such as weather and stadiums. These findings also indicate that more features will be needed to break into the 68% accuracy threshold.

# Conclusion

## Summary of Findings

This research has found that based on the collected historical data and feature set used that weather and the stadium the game takes place have a minimal impact on the outcome of an NFL game.

It should be noted that when the probability threshold was changed from 50/50 to 60/40 the scores of the model improved by a significant margin. The new scores were as follows:

ROC AUC: 0.7028

Brier Score: 0.2183

Accuracy: 0.7389

It should also be noted that due to the change in probability the model only predicted a winner for 1034 games rather than the original 1983 games.

## Limitations

The limitations for this research would be that while there were many more potential features that could have been explored, time constraints and lack of certain types of historical data limited which features could be utilized.

## Future Work

The future work for this research would include exploring other features such as the surface of the playing field and the humidity on the day of the game. The exploration of wind chill and distance a team has had to drive in order to play at a away game could be more features studied in the future. Finally, research could also be done into when teams have their bye week or if they are on a winning or losing streak.

# Refrences

1. Beal, R., Norman, T. J., & Ramchurn, S. D. (2020). A critical comparison of machine learning classifiers to predict match outcomes in the NFL. *International Journal of Computer Science in Sport*, *19*(2), 36–50. https://doi.org/10.2478/ijcss-2020-0009
2. Bosch, P. (2018, February). Predicting the winner of NFL-games using Machine and Deep Learning. . https://cs.vu.nl/~sbhulai/papers/paper-bosch.pdf
3. Bunker, R., & Susnjak, T. (2022, April). The application of machine learning techniques ... - ACM Digital Library. https://dl.acm.org/doi/pdf/10.1613/jair.1.13509
4. M. C. Purucker, "Neural network quarterbacking," in IEEE Potentials, vol. 15, no. 3, pp. 9-15, Aug.-Sept. 1996, doi: 10.1109/45.535226.
5. Warner, J. (2010, December 17). Predicting Margin of Victory in NFL Games: Machine Learning vs. the Las Vegas Line. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=aa120a89092c23127768947c2779f4d937024ce8

**.**