

Predicting the Winner of NFL-games using Elo Rating

Timothy Robbins

Bellevue University

Professor Catie Williams

November 7, 2020

Abstract

Predicting the winner of sporting events has been an exciting topic for decades. Ever since the 1950's, people have attempted to measure the strength of an athlete or team and predict the outcome of such games. One approach I came across, is the use of the Elo rating system, which is basically a measure of the strength of teams in head-to-head games. The website FiveThirtyEight.com uses their own version of the Elo rating system to make predictions that often rival Vegas odds. In this study, I plan to create my own Elo model to predict the winners of NFL games.

The remainder of this study is organized as follows: The background around NFL predictions and the Elo rating system is discussed. I then discuss the methodology, including dataset selection, data preprocessing, and the modeling process. Finally, the study is concluded with the results and conclusion.

Keywords: elo, nfl, American football, Super Bowl, playoffs, team ratings, sports forecasting

Introduction

American football is one of the most popular sports in the United States, and betting on the winner of National Football League (NFL) games is a common practice. Being able to predict the correct winner of an NFL-game can generate significant cash flow. The American Gaming Association (AMA) estimated that fans in the United States spent over \$90 billion on NFL and college games in the 2017 season. The popularity of NFL games has led to the vast release of information on games, players, and statistics. Experts of well-known sports-related organizations, such as ESPN, CBS sports and gambling institutions have made huge investments in the predictions of such games. Numerous attempts have been made to outperform such experts and gambling institutions using data analytics or machine learning.

A popular approach for predicting the outcome of professional sporting events is done using the Elo rating system. An Elo rating system is a method to evaluate teams or players in competitive games. This method was first created and applied to single player games like chess by Arpad Elo, a Hungarian-born American physics professor. Elo ratings were later used to rank and predict team sports, such as the FIFA World Cup, NFL, MLB, NBA, and others. Different sports have found differing levels of predictive power from Elo ratings, with the NFL being one sport where Elo ratings have proved successful.

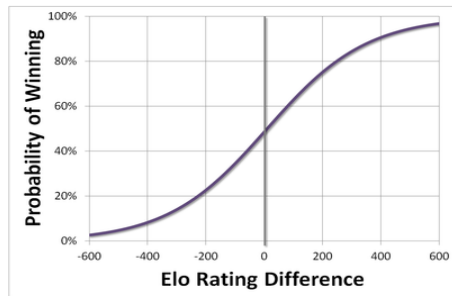
The process of calculating Elo ratings is straightforward. The standard method of implementing an Elo rating system begins with setting every team with a starting value, usually 1,500. All teams start with the same rating. When teams play each other the expected probability of each team winning can be calculated using the following formula:

Exhibit 1 – Home Team Win Probability

$$E_a = \frac{1}{1 + 10^{\frac{R_b - R_a}{400}}}$$

E_a represents the probability of the home team winning, R_b is the Elo rating of the away team and R_a is the Elo rating of the home team prior to the game. The probability of the away team is calculated in the same manner, but since the 2 probabilities must sum to 1 (100%), the away team score can be calculated by subtracting the home team expected probability of winning from 1. The relationship between the ELO difference between the two teams ($R_b - R_a$) and the probability of winning (E_a) can be visualized with the following figure:

Figure 1 – Elo Rating Difference and Probability of Winning



After the game is played, each team's new rating is calculated as the difference between their expected score and their actual score using the following formula:

Exhibit 2 – Updated Elo Rating:

$$R'_A = R_A + K(S_A - E_A).$$

R_A is team A's Elo rating prior to the game, K is the K-factor, S_A represents the actual score, and E_A the probability of team A winning. If the home team wins the game, S_A is reflected as an outcome of 1. If the away team wins S_A is reflected as a 0. The K-factor is a constant term that can be adjusted to determine how much each new game changes a team's Elo rating. The higher the K , the greater the

impact of each individual game on a team's rating. Team B's post-game Elo rating is calculated in the same manner as team A's.

As teams play each other throughout the season, ratings will adjust, and teams will begin to separate themselves based on their performance. Thus, Elo ratings provide a good ranking of each team's performance over time.

Methodology

Datasets

Two different data sets were obtained from FiveThirtyEight.com for this project. First, it was necessary to have starting Elo ratings for each of the 32 teams at the beginning of the 2020 season. FiveThirtyEight is known for making fairly accurate game predictions, so I used the final 2019 Elo ratings of each team as a starting point for the 2020 season, as shown below:

Exhibit 3 – Initial Elo Ratings (1st 5 rows)

team	elo
ARI	1,411.48
ATL	1,549.27
BAL	1,705.15
BUF	1,514.53
CAR	1,374.10

The second dataset obtained from FiveThirtyEight consisted of the team matchups for the 2020 regular season, as well as the actual scores through week 7. There are 16 regular season games for each of the 32 teams, for a total of 512 games. The first 5 records for the dataset are shown below:

Exhibit 4 – 2020 game schedule (1st 5 rows)

date	team1	team2	score1	score2
9/10/2020	KC	HOU	34	20
9/13/2020	JAX	IND	27	20
9/13/2020	DET	CHI	23	27
9/13/2020	NE	MIA	21	11
9/13/2020	CAR	OAK	30	34

After doing some research, I found that team1 represents the home team and team2 the away team. This will be useful later as history has shown the odds to favor a home team win over an away team win. score1 and score 2 represent the actual scores of each game.

Data Preprocessing

The initial datasets I obtained from FiveThirtyEight with the 2020 matchups did not include the week number of the NFL season. Thus, I added this field to each of the datasets by matching up the date of each game with the corresponding week of the 2020 NFL season (there are 16 regular season games for each team, as well as one “bye week” where they do not play).

The second dataset (*Exhibit 4*) contained the scheduled games for the 2020 season, as well as the actual scores of each game through week 7, but the winning team was not indicated. I thus added a simple function to reflect the outcome of each game with 1 indicating a home team win and 0 a home team loss.

I would need to create two key tables to track the results and Elo’s for my project. The first table would be created to track the actual and predicted results of each game, as well as the starting and ending Elo’s of each team by week. The first 6 fields were taken from Dataset 2, and the remaining fields were added:

Exhibit 5 – Game Results Table

Column	Definition
week	Week of the NFL season
date	Date of game
team1	Abbreviation for home team
team2	Abbreviation for away team
score1	Home team's score
score2	Away team's score
actual_outcome	The actual winner of the game
elo1_pre	Home team's Elo rating before the game
elo2_pre	Away team's Elo rating before the game
elo1_prob	Home team's probability of winning according to Elo ratings
elo2_prob	Away team's probability of winning according to Elo ratings
elo1_post	Home team's Elo rating after the game
elo2_post	Away team's Elo rating after the game
pred_outcome	The predicted winner of the game

The second table would be used to keep track of each team's Elo rating throughout each week of the season with the following fields:

Exhibit 6 – Elo Ratings Table

Column	Definition
week	Week of the NFL season
team	Date of game
elo	Each team's final Elo Rating for the week

By resorting this table after each game in descending order by week, I would be able to easily pull the latest Elo rating for each team.

In addition to the two tables I created above, I also created two functions in R: one to calculate each team's winning probability (see *Exhibit 1* above) and another to recalculate each team's Elo rating after each game (see *Exhibit 2* above). As input into the first formula, we take the prior week's Elo rating for the home team and add an additional 100 points to reflect home field advantage and enter this into the model as R_a . Feeding these numbers into the formula to calculate the expected win probability for the home team (E_a) yields a probability between 0 and 1. Since the home team probability of winning and the away team probability of winning must add up to 100%, the away team winning probability (E_b) is determined by subtracting the home team win probability from 1. After the predicted win probabilities are determined, they are compared to the actual game results to calculate the Elo ratings for the next game on their schedule.

Results

In order to evaluate how well my model performed, I calculated a confusion matrix, along with some other useful statistics, as shown below:

Exhibit 7 – Confusion Matrix

Confusion Matrix and Statistics

	Reference	
Prediction	Win	Loss
Win	41	25
Loss	12	27

Accuracy : 0.6476

Sensitivity : 0.7736

Specificity : 0.5192

Pos Pred Value : 0.6212

Neg Pred Value : 0.6923

The confusion matrix shows that there were 41 true positives (win predicted and they won), 25 false positives (win was predicted but they actually lost), 12 false positives (loss predicted but they actually won), and 27 true negatives (loss predicted and they lost). This reflects an accuracy of 64.76% for the 2020 season through week 7, which appears to be pretty accurate, compared to an accuracy of 50% based on guessing.

I also calculated a few other statistics, to shed more light on the predicted vs actual results. Sensitivity, also referred to as recall or the true positive rate (TPR), is a measure of the proportion of correct positive results. The sensitivity of 0.7736 is the rate at which the model correctly predicted wins. On the other hand, specificity, also referred to as the true negative rate (TPR) is a measure is a of correct negative results. The specificity of 0.5192 is the rate at which the model correctly predicted losses. Based on this information, it appears the model is better at predicting the proportion of wins than the proportion of losses.

The positive and negative predictive values (PPV and NPV, respectively) are the proportions of true positive and true negative results, respectively. A high result can be interpreted as indicating the

accuracy of such a statistic. The PPV of 0.6212 indicates that the model is fairly accurate in predicting true positives while the NPV of 0.6923 indicates that the model is even better at predicting true negatives.

Conclusion

The Elo model developed for this study provided results very comparably to FiveThirtyEight's more complex Elo model, which has an accuracy of approximately 65% correct per year. However, it's worth noting that the accuracy of the predictions for the model in this study was based on a relatively small population of 105 games and thus a larger sample could lead to different results.

Future work to predict the outcome of NFL games could include more adjustments to the pre-game Elo ratings, such as the adjusting for the quarterback Elo rating, rest time between games, and distance traveled. In addition, no individual player data was used. Having top players who have gone to the Pro Bowl and/or were awarded MVP could significantly affect the results.

References

- Alvarez, R. (2020). Machine Learning Playoff Predictions: Predicting the Football Greats.
- Balreira, E. C., & Miceli, B. K. (2019). Improving Foresight Predictions in the 2002-2018 NFL Regular-Seasons: A Classic Tale of Quantity vs. Quality. *Journal of Advances in Mathematics and Computer Science*, 1-14.
- Bonds-Raacke, J. M., Fryer, L. S., Nicks, S. D., & Durr, R. T. (2001). Hindsight bias demonstrated in the prediction of a sporting event. *The Journal of social psychology*, 141(3), 349-352.
- Boulier, B. L., & Stekler, H. O. (2003). Predicting the outcomes of National Football League games. *International Journal of forecasting*, 19(2), 257-270.
- Bosch, P. (2018). Predicting the winner of NFL-games using Machine and Deep Learning.
- Harville, D. (1980). Predictions for National Football League games via linear-model methodology. *Journal of the American Statistical Association*, 75(371), 516-524.
- Klein, J., Frowein, A., & Irwin, C. (2018). Predicting Game Day Outcomes in National Football League Games. *SMU Data Science Review*, 1(2), 6.
- Lee, M. D., Danileiko, I., & Vi, J. (2018). Testing the ability of the surprisingly popular method to predict NFL games. *Judgment and Decision Making*, 13(4), 322.
- Miller, T. W. (2015). *Sports analytics and data science: winning the game with methods and models*. FT Press.
- Pelechrinis, K., & Papalexakis, E. (2016). The anatomy of american football: evidence from 7 years of NFL game data. *PLoS one*, 11(12), e0168716.
- Uzoma, A. O., & Nwachukwu, E. O. (2015). A hybrid prediction system for american NFL results. *International Journal of Computer Applications Technology and Research*, 4(01), 42-47.
- how-our-nfl-predictions-work. (2020, October 15). Retrieved from FiveThirtyEight.com: <https://fivethirtyeight.com/methodology/how-our-nfl-predictions-work/>

Appendix

The illustration below shows the change in each teams' Elo rating through week 7 of the NFL season by division.

