NFL Sports Betting: Strategies, Optimization, and Automation

Travis Nitkiewicz, Jackson Ewers, Nadav Langberg

April 2023

1 Abstract

With the rise in popularity of sports betting thanks to the advent of platforms such as DraftKings and FanDuel, just about any fan over the age of twenty-one can gamble on their favorite NFL teams. We intend to uncover several profitable strategies using items such as points spread, moneyline, over/under, and others to identify methods to reliably and consistently profit.

In this project, we will discuss a novel, mathematical approach to bet on a team, where the metric that evaluates the strength of our algorithm is increased wealth level above emotional betting and the expected random betting outcome.

Methodologies and tools used during the process are excel, r, and chatgpt. We intend to implement a test train split of our data to backtest it on previous years, and use other mathematical tools such as various machine learning and regression strategies, the Kelly criterion, and statistical t tests to seek an accurate, profitable outcome within sports betting.

The project is split into two sections; the first section investigates NFL betting on season win totals, using the previous year of data to predict the next year's results. The second section explores NFL betting week-to-week, using the first few weeks of the season to calibrate the model. Some methods used perform significantly better than others, especially when testing various machine learning methods on the data. Additionally, different methods work well for different sports and betting strategies.

As a full disclaimer, this project is an academic research project. It is not intended to be used in betting and not financial advice. These are simply our personal opinions on the topic and are not representative of Michigan State University or the Mathematics Department.

2 NFL Prediction of Success Season over Season

2.1 Introduction

In exploring the question "how can a data-driven statistical approach identify profitable weekly betting outcomes in the NFL", we wanted to explore predicting NFL outcomes year-over-year. In order to do this, the process was broken up into three steps. Step one, source data from a variety of websites and platforms over the years 2014-2018. Step two, identify the important variables to team success. Step three, using those variables, predict, utilizing a variety of statistical and machine learning tools, how successful the teams will be in the next year. Then, use these predictions to attempt to have a profitable betting system on the upcoming year total team wins betting lines. The tools used throughout the process were excel for preliminary analysis and organization, R for data analysis, and chatgpt/ stackexchange/ geeksforgeeks for debugging. The methods used throughout were linear regression, lasso regression, logistic regression, best subset selection, binomial regression, and support vector machines. Lastly, the Kelly Criterion was calculated to determine the optimal betting amount.

2.2 Process

2.2.1 Data Sourcing and Collection

The data was sourced from a variety of websites and databases. The first set of data collected was from the mathletics github repository, and provided numerous metrics on each NFL team. The variables mathletics provides are as follows [7]:

- Year, the year of the NFL team.
- Team, the team name.
- Margin, points margin over the course of the season.
- RET TD, the total number of returned touchdowns from a kickoff/punt.
- PENDIF, the total of penalty yards for vs against the team in each game, summed over the course of the season.
- PY/A, the number of passing yards per attempt.
- DPY/A, the number of passing yards the defense allowed per attempt.
- RY/A, the number of rushing yards per attempt.
- DRY/A, the number of rushing yards the defense allowed per attempt.
- TO, the number of offensive turnovers, interceptions and fumbles but not 4th down stops.
- DTO, the number of turnovers the defense earned. Includes interceptions and fumbles but not 4th down stops.

The next round of data was sourced from Pro-Football-Reference [6].

- Wins, the number of games won in a season.
- Losses, the number of games lost in a season.

- SoS, the strength of schedule in a season, 0 is the average, a positive number means a tougher schedule.
- Playoffs, if the team made the playoffs that year.
- Superbowl, if the team won the superbowl that year.

The next group of data was utilized from BetIQ, a sports betting trends site [1].

- O/U Games, a variable that shows the sports betting odds for how many games the team is expected to win in that season.
- OverLine, the payoff if the team hits the over.
- UnderLine, the payoff if the team hits the under.
- OverPerc, the number of times the team hit the over on points in each game that season.
- UnderPerc, the number of times the team hit the under on points in each game that season.
- ATS, Against the Spread, the average amount of points that the team goes over the expected points.

The last round of sourced data came from wikipedia, simply on the types of NFL stadiums [8].

- Stadium Type, if the stadium is an indoor, outdoor, or retractable stadium.
- Grass Type, if the stadium has grass or turf.

Lastly, the remaining variables are calculations from the Pro-Football-Reference data [6], but they were computed by us.

- Division Wins, how many wins the rest of the teams in the division had that year.
- Division Losses, how many losses the other teams in the division had that year.
- StdDevDivWins, the standard deviation of the division wins, examines how spread out the division was.
- Worst in Division, a categorical variable indicating if the team had the worst record in the division. If there was a tie, there was no worst team listed.
- Best in Division, same as Worst in Division but for the teams with the best record.

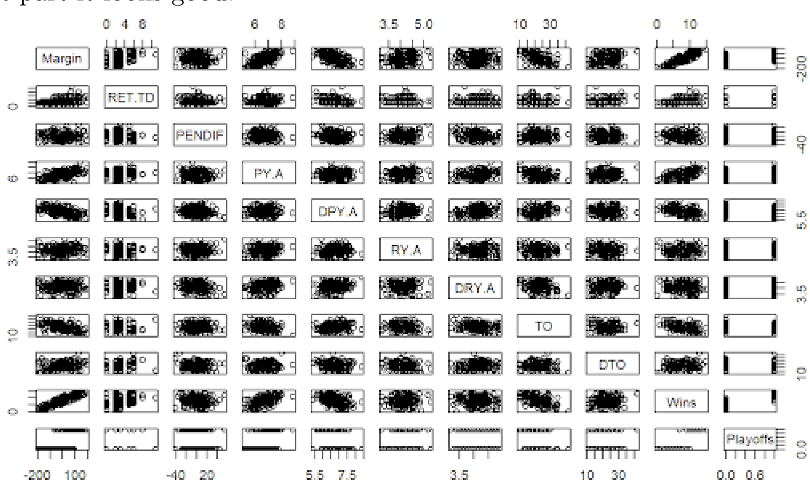
- Games Ahead Last, how many games ahead of the last place division team the current team is.
- Game Behind First, how many games behind first in the division the current team is

Below are two snippets of the dataset. The data ranges from 2014-17, and has all of the variables listed above.

Year	Team	Margin	RET TD	PENDIF	PY/A	DPY/A	RY/A	DRY/A	TO	DTO	Wins	Losses	SoS	O/U Games	OverLine	UnderLine	Playoffs	SuperBowl
2017	Arizona	-68	2	-20	6.1	6.3	3.4	3.5	26	21	8	8	0.4	8.5	-150	120	No	No
2017	Atlanta	40	2	-1	7.4	6.2	4.2	4.2	16	18	10	6	1.9	9.5	-125	-105	Yes	No
2017	Baltimore	93	6	-5	5.3	6	4	4.1	18	34	9	7	-2.4	8	-115	115	No	No
2017	Buffalo	-61	3	2	5.7	6.3	4.1	4.3	18	24	9	7	-0.5	6	-155	125	Yes	No
2017	Carolina	28	3	-22	6.2	6.9	4.3	3.9	21	21	11	5	2.1	8.5	-180	150	Yes	No
2017	Chicago	-56	5	-3	5.9	6.4	4.2	4	22	22	5	11	2.2	5.5	-100	-130	No	No
2017	Cincinnati	-59	3	4	6.1	6	3.6	4.2	22	14	7	9	-1.3	8.5	-100	-130	No	No
2017	Cleveland	-176	2	9	5.6	7	4.5	3.4	42	13	0	16	0	5	-105	125	No	No
2017	Dallas	21	2	-9	6.4	6.1	4.5	4.1	22	21	9	7	0.2	9.5	-100	-130	No	No
2017	Denver	-92	3	24	5.9	6.5	4.1	3.3	34	18	5	11	-0.9	8.5	125	-155	No	No

2.2.2 Identifying Factors Key to Team Success in a Season

The initial step was to check for correlations between the variables to avoid multicollinearity problems. Unfortunately, there are so many variables, checking all of them is challenging and unproductive. So, we checked primarily the numeric variables and the ones we anticipated would be most impactful. Below is the correlation plot. As you can see, there are a few strong correlations, such as winning and having a higher point margin than your opponent, but for the most part it looks good.



After sourcing the data and checking for correlation, the next step is to figure out what factors drive team success. Using this data, there are five ways to measure team success: total wins, made playoffs, won the superbowl, won the division, and points margin. To identify these factors, we conducted a simple linear regression using all the variables (aside from these five and counterparts like number of losses) and used each factor as the response.

After conducting the initial analysis, we removed SoS and StdDevDivWins, as they were the most insignificant (p value greater than 0.5), to help pair down the analysis, and tried again.

After running a more streamlined analysis, we then removed another set of predictors based on significance again, leaving us with just Return TDs, Penalty Difference, PYA, RYA, DPYA, DRYA, TO, and DTO. Running linear regression on just these variables, across all 5 response variables, yielded some interesting results. Using a 10 percent significance level, we summed how many times each variable was significant. The results are as follows:

	Ret TD	PENDIF	PYA	RYA	DPYA	DRYA	TO	DTO	StdDevDiv
Margin	1	0	1	0	1	1	1	1	0
Wins	1	1	1	0	1	1	1	1	0
Playoffs	0	1	1	0	1	1	1	0	0
Superbowl	1	0	0	0	0	1	1	0	0
Division	0	0	1	0	1	0	1	0	1
Totals	**3**	**2**	**4**	**0**	**4**	**4**	**5**	**2**	**1**

Looking at the results, TO (Turnovers) is the most important across all types of success, followed by PYA, DPYA, and DRYA. So, the saying, "defensive wins championships", appears to have some merit, as both defensive metrics were significant in all but one model. In the predictions to follow, these metrics will primarily be used.

2.2.3 Prediction Year Over Year

The process for each method is relatively similar. Each method splits the data using a 70-30 test train split. This was chosen since there are only 32 teams, so a higher split would have very few test predictions. All models use 2014 as the training data, and added to that data is the betting line information for 2015 (Over/Under number of games won, the over line, and the under line), since that comes out well before the 2015 season. The models test on the 2015 season wins, and compares both the model predicted wins and the actual season wins to the 2015 over/under line, and converts these into a betting option: over or under. The model then tests the accuracy of these predictions over 100 random seeds, to ensure the sample of 10 test points is sufficiently representative. A data frame of the results of one seed from linear regression is included to better show the process:

	Year	Team	Last_Year_Wins	Pred_Wins	act_wins	OU_Line	OU_pred	OU_act
98	2015	Atlanta	6	2.921338	8	8.5	Under	Under
101	2015	Carolina	7	8.107399	15	8.5	Under	Over
106	2015	Denver	12	11.339709	12	10.5	Over	Over
109	2015	Houston	9	5.497676	9	8.5	Under	Over
110	2015	Indianapolis	11	4.567737	8	10.5	Under	Under
115	2015	Miami	8	8.306038	6	8.5	Under	Under
116	2015	Minnesota	7	7.902985	11	7.0	Over	Over
121	2015	Oakland	3	10.936177	7	5.5	Over	Over
122	2015	Philadelphia	10	9.248784	7	9.5	Under	Under
127	2015	Tennessee	2	6.029309	3	5.5	Over	Under

1-10 of 10 rows

In this dataframe, ActWins is the number of wins in the next season, the

2015 season, and the OU-Line is the over/under line for total wins in a season. So in this particular seed, the model predicts (OU-pred column) "Under" 6 times and "Over" 4 times. The accuracy is derived from comparing the columns OU-pred, the model prediction, and OU-act, how the actual next season wins fared against the betting lines. In this sample, the model is correct 7/10 times, only failing for Carolina, Tennessee, and Houston. All the models use 2014 data as training data and 2015 data as testing data, and then the most successful models explore 2016 and 2017 data.

The first type of analysis conducted was multiple linear regression. Initially, we used all predictor variables aside from losses to predict wins. We did this on just the 2014 data to start off with. The adjusted R^2 value indicated that 98 percent of the variability in wins was explained by our data, which makes sense given we used all predictors. However, very few predictors were statistically significant even at the 10 percent level, meaning that it is hard to differentiate between if the predictor variables are impacting wins or if it is just random. The model with all predictors yielded a 45 percent accuracy. To attempt to improve predictions, we decided to remove several variables because so many were not statistically significant.

In order to better identify fewer, more important predictors, we used best subset selection from the leaps package. We chose the method best subset instead of forwards or backwards stepwise selection because the dataset is relatively small (only 4 years of data x 32 teams) so we were not worried about the computational costs that can derail best subset selection. After performing best subset selection, we found the top predictors were Margin, RY.A, DRY.A, SoS, Playoffs, and UnderLine for the next season. Applying these variables to a linear regression model unfortunately yielded an accuracy rating of 41 percent across 100 simulations of the test train split. With further experimentation on which variables to include, the accuracy dropped to 39

Since linear regression and best subset selection were not predicting above 53 percent, the next approach we tried was regularization. We chose lasso regression instead of ridge regression since we are trying to eliminate some of the insignificant variables from our regression, and lasso reduces insignificant coefficients to zero. After completing lasso regression using the glmnet package, the model used SoS, O.U.Games, UnderLine for 2014, ATS, BestInDivision, GamesAheadLast, and O.U.Games for the next season as the important predictors. The best lambda parameter value was -0.4, computed by cross validation. This model had an average accuracy of 51.7 percent, which is good, but short of the goal of hitting at least 53 percent accuracy.

After working with regression for so much of the data, we decided to flip to classification and implement logistic regression. Since with logistic regression you have two choices, we chose to predict having a winning or losing season. If it is promising, logistic regression could be used to predict making playoffs or predicting thresholds of wins. Unfortunately, logistic regression yielded an accuracy of 43.1 percent across 100 simulations, and since that was so far below the goal of 53 percent we decided to not pursue this idea further. To best illustrate logistic regression, the results were put in a table form:

	Year	Team	Last_Year_Win_Prop	Pred_Win_Prop	act_wins_Prop	OU_Line_Prop	OU_Pred	OU_Act
97	2015	Buffalo	0.5625	0.2889189	0.5000	0.53125	Winning_Season	Tie
98	2015	Carolina	0.4375	0.3617386	0.9375	0.53125	Winning_Season	Winning_Season
101	2015	Cincinnati	0.6250	0.5399916	0.7500	0.53125	Winning_Season	Winning_Season
105	2015	Cleveland	0.4375	0.6857402	0.1875	0.40625	Losing_Season	Losing_Season
106	2015	Dallas	0.7500	0.5080232	0.2500	0.59375	Winning_Season	Losing_Season
112	2015	Denver	0.7500	0.4616012	0.7500	0.65625	Winning_Season	Winning_Season
120	2015	Indianapolis	0.6875	0.3520234	0.5000	0.65625	Winning_Season	Tie
121	2015	Jacksonville	0.1875	0.4409228	0.3125	0.34375	Losing_Season	Losing_Season
122	2015	Seattle	0.7500	0.4953221	0.6250	0.68750	Winning_Season	Winning_Season
126	2015	Tennessee	0.1250	0.2923610	0.1875	0.34375	Losing_Season	Losing_Season

Every number is a proportion of total season wins. The OU-Pred column is the column that translates the oddsmakers line into a prediction; for example, predicting 7 wins would be predicting a losing season. This whole concept would be better suited to using playoffs or some other categorical variable as the response, but since the accuracy was fairly low, we did not explore further.

Binomial regression was suggested by several professors as a better way to run regression on data like wins, where each event is independent and there are two outcomes, in this case win or lose. After implementing binomial regression using all the predictors, the accuracy was 47 percent. With some adjustments to the variables, the accuracy climbed to 51 percent, which is good. However, as you can see in the model below, none of the variables used are anywhere close to being statistically significant, with the best p value in the last column being 0.297, very far from typical significance standards of 0.05 or 0.10.

```
[1] 51
The overall success percentage is  0.51  and the average mse is  63.7931NULL
The average error rate is  0.5  and the average precision of the model is  0.4958095NULL

Call:
lm(formula = Wins_15/16 ~ PY.A + DPY.A + DRY.A + TO + OverLine_15/+O.U.Games_15 +
    UnderLine_15, data = train_data_14)

Residuals:
     Min       1Q   Median       3Q      Max
-0.33294 -0.03268  0.00403  0.08716  0.27825

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                0.6207897  0.8240211   0.753    0.464
PY.A                       0.0430676  0.0802454   0.537    0.600
DPY.A                     -0.0241345  0.0804134  -0.300    0.768
DRY.A                     -0.0229619  0.0959960  -0.239    0.814
TO                        -0.0095706  0.0088261  -1.084    0.297
OverLine_15                0.0008313  0.0020562   0.404    0.692
UnderLine_15              -0.0004270  0.0005191  -0.823    0.425
OverLine_15:O.U.Games_15 -0.0001226  0.0002335  -0.525    0.608
```

Support Vector Machines (SVMs) returned the highest accuracy out of any prediction method. SVMs take the data, move the data to a higher dimensionality, and split the data along a hyperplane that minimizes area of the residuals. When testing on the 2014-15 data, predictions yielded in the range of 50-65 percent. Adding more predictors to the model generally increased accuracy on the current but also will likely increase overfitting, as accuracy often dropped in the next year. Only the radial kernel performed well; linear, polynomial, and sigmoid kernels all yielded around or below 50 percent accuracy. Since SVMs had such a higher accuracy compared to other methods, we chose to apply only SVMs to the rest of the data.

To go through and identify the optimal SVM, we tested five SVM models. The low predictors model, with only 6 predictors, and then for each set of data (14-15, 15-16, 16-17), we tuned the predictors and parameters to whatever yielded the best accuracy. So, there are four models, referred to as the few params, 14-15 (best) model, 15-16 model, and the 16-17 model. Since the 16-17 model appeared to be performing well, we also made a 16-17 model with fewer variables due to concerns with overfitting. To choose the predictor variables, this was simply manual trial and error for best accuracy when predicting. Within each model, there is a cost parameter that we changed to try to help with overfitting and to increase the accuracy. The results have been organized into the table below:

Model Name	Few Params	Few Params	Few Params	Few Params	14-15 model	14-15 model	15-16 model	15-16 model	16-17 model	16-17 model	16-17 model	16-17 mod few pred	16-17 mod few pred	16-17 mod few pred
Model Cost	0.4	0.5	1	3	1	3	1	1.6	0.5	1	1.5	0.5	1	1.5
Data_14_15	54.4	54.4	54.6	52.4	63	65.1	53.8	56.7	58.7	60.7	59.6	56.9	56.9	56.5
Data_15_16	59.2	60.4	56	54.3	47.6	42.8	58.2	61.1	58	55.3	54.3	58.5	56.8	55.7
Data_16_17	60.6	61	57	54.7	63.2	58.5	63.8	61.5	67.9	68.9	68.4	68.1	66.9	64.8
Avg	**58.07**	**58.60**	**55.87**	**53.80**	**57.93**	**55.47**	**58.60**	**59.77**	**61.53**	**61.63**	**60.77**	**61.17**	**60.20**	**59.00**

To quickly explain the table, the columns represent each model tested. So for example, the first model tested is the 'few parameters' model with a cost parameter of 0.4. Using the first column, the accuracy rating of this 'few params, cost = 0.4' model on the 2014-15 data, 15-16 data, and 16-17 data is listed. Then, the last cell in the first column is the average accuracy rating across the three data splits. In dark green represents the best performing model for that data split, red being the worst, and light green being the next top four performing.

Some key takeaways from the table:

- The 16-17 model had the highest average accuracy across the models, with 60-61 percent depending on the cost parameter.
- The 16-17 model with cost = 0.5 had a very good average (61.5 percent) as well as high levels across the board (low score of 58.0
- Lower cost parameters yielded higher accuracy across three out of four models The 14-15 model performs quite poorly on the next year of data. However, the 15-16 model returns high accuracy (61-63 percent) on the next year of data
- The few parameters model performs very consistently (54.4-60), but lower net average accuracy
- The model 16-17 with cost = 0.5 appears to be the best combination of higher accuracy and consistency
- The model 16-17 with fewer parameters also performs quite well, and may be useful when dealing with overfitting.

2.3 Results - Test Data Accuracy, Kelly Criterion, and Betting

Now there is an optimal model, the 16-17 model, both regular and with fewer predictors, at cost =0.5. When testing this model on the test data, the 16-17 model does not do as well, with an accuracy on the 2018 test data of 0.484. However, when using the fewer predictors 16-17 model, the accuracy improves to 0.548, likely due to overfitting differences. Note that Washington has been removed from the test data since that game was a push, meaning Washington got exactly the same number of wins as the OU-line wagered for, meaning bettors simply get their money back.

Now that we have an accuracy for the model, we can calculate the Kelly Criterion. The Kelly Criterion is a method for computing the best amount of your portfolio to wager on each bet. The Kelly Criterion is computed via the formula f = [bp - q] / b, where f is the fraction of your portfolio you bet, b is the odds on the bet, p is probability of winning, and q is probability of losing. Due to concerns of potentially overfitting, we will use accuracy = 0.6117, the average of the fewer predictors model,0.569, the lowest training score of the 16-17 fewer predictors model, and 0.548, the true accuracy, to capture a range of outcomes. So b is going to be different for each bet, but p and q will be 0.6117, 0.569, or 0.548, and then q = 1-p,. To calculate b, first convert from the sportsbook (ex -110), to odds. To do this, one must do 100/absolute value ([line]) + 1. An example, odds of -110, yields 100/110 as odds and then b = 100 / 110 = 0.909.

However, the goal of this betting operation is to bet on most/all nfl games, since the model is spread across all games. The problem with the Kelly Criterion is that it often suggests a higher percentage wager, which given there are 32 bets the model should make, it is likely the Kelly Criterion will suggest percentages that exceed 100 percent in total. So, we will first test a model where we just wager equally on each game. After that, we will test a model that utilizes a weighted Kelly Criterion approach.

The weighted Kelly Criterion approach consists of computing the k value for each bet. Then, take the average of all of those k values. Then make a weighted k score for each team/bet, where the weighted score is the k value divided by the average k value. This means that if the Kelly criterion is suggesting to bet 20 percent, when the average is 15 percent, you would bet 20/15, or 1.25 units on that bet. This is an approach that allows to weight bets the Kelly Criterion rates as higher potential paying as more. Additionally, the Kelly Criterion suggests that if the k value is negative, to not bet on that team as the outcome is expected to lose money.

The results from the evenly distributed betting and the weighted Kelly Criterion betting are organized below. It is assumed a portfolio of 3200 dollars, so that one could bet 100 dollars on each team. Some other investment metrics are included for comparison

	Unit Size	Dollar Return	Percent Return
Equal Weight	1, $100	$184.94	5.78%
Accuracy 0.6117	1, $100	$350.59	10.96%
Accuracy 0.569	1.95, $195	$542.83	16.96%
Accuracy 0.548	3.49, $349	$837.21	26.16%
S and P 500 average annual return 2014-18	$3200 invested	$339.52	10.61%
US 1-Year Treasury Average annual return 14-18	$3200 invested	$29.25	0.914%
US 10-Year Treasury Average annual return 14-18	$3200 invested	$75.26	2.35%
Average Annual Inflation 2014-18	$3200 cash	-$48.64	-1.52%

Looking at the results, it is clear following the Kelly Criterion significantly improves the outcome. However, this only works if you follow the weighting as well, as the powerful return on the lowest accuracy, with a unit size of 1, would just be 26.16 percent / 3.49 = 7.50 percent. Of course, when betting with a unit size that high (349 dollars on a portfolio of 3200 dollars is just over 10 percent capital per bet), having an accuracy above 0.53 becomes critically important. If bettors are uncomfortable risking that large of a position size, it will likely preserve capital and reduce risk at the cost of lower returns. However, looking at these results, it is indeed possible that one can predict and profit from NFL win total odds for the next season using the prior season's data as training data

2.4 Further Steps

Despite the breadth of methods implemented on this data, there are still many more ways to predict. To name a few, generalized additive models with natural or smoothing splines, polynomial regression, decision trees, bagging, boosting, principal component analysis and regression, quadratic or linear discriminant analysis, and neural networks are all machine learning techniques that could be further explored to find a potential higher accuracy. Additionally, applying different transformations to the data, such as log transformations or exponential, investigation into removal of outliers, and combining multiple years of training data are all other ways to manipulate the dataset to potentially improve prediction.

On the betting side, exploring how wagering differently on higher magnitude predictions could potentially increase profitability. Furthermore, investigating more into how prediction accuracy plays into total return could potentially bear fruit, as the model did best when using the true, and the lowest, accuracy.

On the general NFL side, applying this approach to predicting using classification, on which teams will make the playoffs or win their division would be interesting to see. And, applying this methodology to a week to week setting would be fascinating, as there should be less variation in week-to-week NFL data than season-over-season data. Lastly, to better predict how teams will perform in the next year, looking into the creation of a player/coach relative value index would be interesting to investigate. If a high value player or coach moves teams, that definitely has an impact on the next season, and one that this model can only capture through the next year betting lines. An index would help capture more information about offseason changes.

2.5 Takeaways

In exploring the question "how can a novel mathematical approach identify profitable weekly betting outcomes in the NFL", we wanted to explore predicting NFL outcomes year-over-year. First, the factors that were most important to team success were offensive turnovers, defensive rushing yards per attempt, defensive passing yards per attempt, and passing yards per attempt. Using these variables as our base for analysis, we attempted to predict for every NFL team if they would hit over or under the betting line for wins in the season. We explored linear regression, best subset selection, lasso regression, logistic regression, binomial regression, and support vector machines to try and predict above 53 percent, the generally accepted profitable winrate in sports betting.

The results were interesting; support vector machines worked the best by far, followed by regularization/shrinkage using the lasso method and binomial regression. These were the only methods that had accuracy scores of over 50 percent. Using support vector machines, we were able to find a combination of variables and parameters that yielded an average of 61 percent accuracy, using the 2014, 15, and 16 data as training data on the following years. Applying the model to the 2018 data for testing, the model had an accuracy of 54.8 percent, indicating some signs of overfitting, but still enough to be profitable.

Applying the model to betting, we compared a portfolio of bets where each one was equally weighted, and another approach where, using the Kelly Criterion, bets were maximized based on expected payoff. The Kelly Criterion method was more successful every time, regardless of the prediction accuracy value used. The betting performed best with the accuracy parameter representing the true accuracy, but the returns ranged from 5.78 percent to 26.16 percent depending on the strategy. In conclusion, it is indeed possible to predict with some degree of accuracy and profit from these predictions in the NFL season over season data when looking at the 2014-18 sample set. We hope similar methods across more recent data will yield comparable results, but with the legalization of sports betting in the US in 2018, this change may affect future results.

3 Optimal Betting Using an adjusted Odds Ratio for Spreads and Over/Unders

3.1 Introduction

Historically, individuals used past trends and team statistics to develop an educated bet on a particular team. In addition to the observed trends, many bettors used emotions and illogical thinking to influence their bets.

We define emotional betting as mindless parley making (multiple legs for astronomical odds) or placing bets based on feelings, in conjunction with minimal research (For example - Allen will throw for over 225 yards this weekend because hes averaging over 225 yards/ game for the season. But this logic isn't fully accurate because it doesn't take opposing defenses into account.)

In this paper, we will discuss a novel, mathematical approach to bet on a team, where the metric that evaluates the strength of our algorithm is increased wealth level above emotional betting and the expected random betting outcome.

Our initial mathematical approach (when tested in real time) provides a betting strategy as effective, if not more effective than the classic historical trends.

Furthermore, the betting strategy can be optimized further using the Kelly criterion. We present our results as well as statistical testing of the observed data and well use a t-test to validate our statistical data

3.2 Review of Established Prediction and Evaluation Metrics in Sports

3.2.1 Pythagorean Theorem of Baseball

One of the earliest approaches to link in-game performance to season long statistics is the Pythagorean Theorem of Baseball [3] [5] [7], where the relevant variables are the ratio R of points scored for vs points scored against:

$$R = \frac{PointsFor}{PointsAgainst} \tag{1}$$

and the corresponding prediction for win percentage for the season (listed as a decimal between 0 and 1) based on (1)

$$PToB = \frac{R^n}{R^n + 1}. \tag{2}$$

Using historical league data for a teams performance, one can calibrate an optimal n to predict every teams record by minimizing an appropriate loss functional. A negative deviation leads to the assumption that the team is "undervalued" - meaning the teams better than its record suggests. Contrarily, a positive deviation leads to a team being "overvalued". These valuations come into play when choosing which team to bet on in any particular game, and the theorem suggests that an undervalued team will over-perform in the playoffs, when compared to an overvalued team. [7]

3.3 Evolution of formula

An initial approach was to use the 2022 NFL playoffs results as input data to the Pythagorean algorithm. Every team was marked as undervalued, slightly undervalued, slightly overvalued, or overvalued. Using these classifications, we decided that a team would win and cover the spread if they were more undervalued than the team they are facing. Our initial approach had more success in predicting the spread than the winner. This led us to believe that there's a deeper discrepancy between beating the spread and winning. Naturally, this led to a discussion about team quality, an undervalued team isn't necessarily a better team, even in a playoff scenario.

This opens up our analysis to link team quality with win prediction. Our first approach is to measure efficiency via a team's kickoff/ punt ratio. As a simplification, we ignore turnovers on downs or fumbles/ interceptions and assume that every football drive can end in one of two ways, a kickoff (following a score) or a punt. This ratio is a multiplicative factor to enhance (1) and account for drive success. This enhanced ratio is what we subsequently refer to as our *magic number*. In addition to this magic number, which was used to predict the winner, we also track if a team covers the spread and display that in our tables below using various colors.

$$MagicNumber := RR = R * \frac{Kickoffs}{Punts} \tag{3}$$

Below are the results from the 2021-2022 seasons playoff match ups. The color column show which team had the more undervalued color, RR is what we refer to as the magic number and has the team that has the higher ratio in that specific game. ATS (against-the-spread) and ML (money-line) are our results columns, where a green in the ATS column means that the team in the color column was able to cover, and a green in the ML column means that a team with the higher magic number (or RR) won the game outright. For both these columns, a red represents a loss or a team not covering. From this analysis of the 2022 playoffs, this early strategy successfully predicted the spread 61.5 percent of the time, and predicted 69 percent of the winners.

Color	Ratio*ratio	ATS	ML
NA	BILLS		
Bengals	BAN		
KC	KC		
SF	DAL		
Eagles	TB		
cardinals	LAR		
bengals	TEN		
SF	GB		
KC	KC		
NA	LAR		
bengals	KC		
49ers	LAR		
bengals	LAR		
0.5	anomoly	8 green	9 green
5	normal	5 red	4 red

As mentioned above, this is clearly a simplification, drives can end result in fumbles, interceptions, or other factors. As the formula evolved, we've included those factors in other metrics. To include more predictive factors and increase the accuracy of the formula, we added three new main components, the defensive number, the offensive number, and SOS (strength-of-schedule).

3.3.1 Defensive Rating

The defensive number was added first, the original magic number seemed to be very offense-centered, teams with a low offensive rating but a higher defensive rating were undervalued and weren't considered high-caliber teams. The defensive rating takes many key defensive metrics (using metrics that'll show its strength and weakness), and is then added to the RR to give a more accurate rating.

3.3.2 Offensive Rating

The addition of the defensive rating led to an overcorrection in the ratings of each team, many defensive-centered teams were overvalued. To fix this, an offensive rating was created. The offensive rating is composed of many key metrics that imply how strong an offense is (for example, the Eagles and the Chiefs have the two highest offensive ratings, while a team like the Broncos has a very low offensive rating). The offensive rating was added to the magic number to create a more accurate rating.

3.3.3 Strength of Schedule

For a final adjustment, we adjust a team's rating for the difficulty of the opponents. Not all team schedules are equivalent in difficulty. To accommodate for this imbalance, the magic number is multiplied by the SOS. SOS is the average record (from the season prior) of the teams that a team is going to play against. [2]

After all of these base adjustments, minor adjustments were made with the weights of the defensive rating, offensive rating and RR to fix any teams that were overvalued or undervalued.

These adjustments lead us to the most recent iteration of the magic number:

$$\bar{R} = (0.7RR + 0.5DEF + 0.5OFF) * SOS \tag{4}$$

Later in the process, we decided to use the offensive and defensive ratings to predict the Over/Under of a game. To perform this effectively, we adjusted the defensive and offensive ratings to be centered around 1.0. If a game's combined offensive rating is higher than its defensive rating, the score will be over the predicted score, and if the opposite happens, then the match score will be under the predicted value.

Tm	defensive ratio	DEF R OVUN	OFF ratio	OFF R OVUN
Arizona Cardinals	0.689	0.630	0.627	1.010
Atlanta Falcons	0.691	0.669	0.563	0.829
Baltimore Ravens*	1.540	1.382	0.690	1.064
Buffalo Bills*	1.543	1.425	0.686	0.957
Carolina Panthers	0.875	0.770	0.467	0.686
Chicago Bears	0.622	0.606	0.526	0.843
Cincinnati Bengals*	1.068	1.007	0.881	1.515
Cleveland Browns	0.742	0.653	0.645	0.929
Dallas Cowboys+	2.250	1.949	1.175	1.640
Denver Broncos	1.750	1.513	0.326	0.569
Detroit Lions	0.703	0.654	0.987	1.398
Green Bay Packers	0.818	0.752	0.652	0.931
Houston Texans	1.120	0.992	0.319	0.470
Indianapolis Colts	0.903	0.762	0.267	0.399
Jacksonville Jaguars	0.823	0.791	0.675	0.980
Kansas City Chiefs*	0.968	0.800	0.988	1.367
Las Vegas Raiders	0.643	0.548	0.827	1.238
Los Angeles Chargers	0.848	0.805	0.709	1.102
Los Angeles Rams	1.038	0.931	0.365	0.583
Miami Dolphins+	0.828	0.710	0.767	1.144
Minnesota Vikings*	1.207	1.120	0.821	1.339
New England Patriots	1.750	1.543	0.420	0.638
New Orleans Saints	0.963	0.765	0.518	0.731
New York Giants+	1.100	1.004	0.612	1.061
New York Jets+	1.435	1.232	0.575	0.864
Philadelphia Eagles*	1.792	1.599	1.229	2.113
Pittsburgh Steelers	0.950	0.896	0.409	0.652
San Francisco 49ers*	1.800	1.594	0.734	1.082
Seattle Seahawks+	1.088	1.007	0.687	1.036
Tampa Bay Buccaneers*	1.200	0.968	0.645	0.950
Tennessee Titans*	1.093	0.929	0.613	1.035
Washington Commanders	1.185	1.028	0.536	0.872
		1.001	0.654	1.001

3.4 Bet Size and Weight

It appears that equation (4) can give us an edge over traditional sports books, and we further enhance this approach by incorporating the Kelly criterion to optimize our bet size and weight.

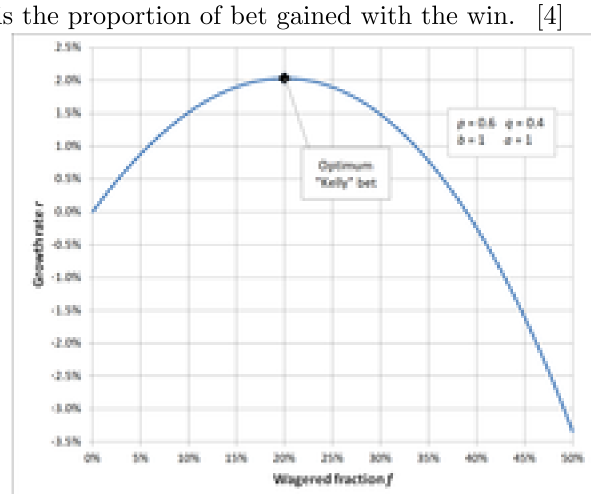
3.4.1 Kelly Criterion: Overview and implementation-Overview

The Kelly Criterion is a gambling/investment strategy that determines the theoretical optimal size for a bet or asset investment. It is the result of maximizing the expected value of the logarithm of future wealth resulting from investment.

The image below is a good example of returns under the Kelly criterion in comparison to other rates. The formula is outlined as such:

$$F^* = p - \frac{1-p}{b} \tag{5}$$

Where F is the fraction of the current bankroll. p is the probability of a win. b is the proportion of bet gained with the win. [4]



[9]

3.4.2 Kelly Criterion for our formula

Using 8 Weeks of empirical evidence, we were able to get stable rates to use for the Kelly criterion. To simplify some factors were going to assume the odds of the spread and over/under is -110 (meaning you need to bet 110 dollars to win 100). Spreads will usually range from -115 to -105 normally as people bet more on a particular team during the week.

With this main assumption, our variables have the following value: $p = 0.6415$ and $b = 0.909$. Plugging these into the formula gives us $F^* = 0.247$. This in turn would give us a geometric growth rate of 1.039 per bet. To summarize, we should theoretically bet around 24.7 percent of our pot each betting round. Due to how close football games are in time slots, we've decided that instead of betting 24.7 percent on each game, well bet 24.7 percent on each time-slot, giving prime-time games a higher weight when compared to the 1pm slot of games.

3.5 Empirical findings

After analyzing a multitude of NFL weeks, compiling data and using the data to predict the correct bets, we've compiled the following results.

Summery after week 13:

OV/UN	Spread	ML gain	
60	53	-725	
0.674157303	0.609195402		
1.301652893	1.149793388	0.902027027	
off rating - def rating	Mod magic number	bet on underdog	

3.5.1 OV/UN

The first column in the graph shows the success rate of the Over/Under (OV/UN) bets and the ratio of money earned/ money bet.

As seen above, the success rate of OV/UN is 0.67, meaning that every bet we've placed as a 67 percent chance of being correct. While this success rate seems much higher than the expected (the expected rate is 50 percent), it also needs to be proven to be statistically significant. To prove that, we will perform a basic t-test.

For a t value to be significant, it has to be greater than 2 (at the 95 percent confidence level).

$$t = \frac{\bar{x} - \mu}{SD/\sqrt{n}} \tag{6}$$

Where:
$\bar{x}$ is the experimental mean
μ is the expected mean
SD is the standard deviation of the sample
and n is the number of data points in the sample

For the OV/UN; x = 0.674, μ = .5, the $SD = \sqrt{\frac{\sigma^2}{n-1}}$, where σ^2 is the variance of a Bernoulli variable with p = .5, and n = 89.

Using these numbers, the calculated t value is 6.415, meaning the findings are statistically significant.

3.5.2 Spreads

Next, the second column in the graph shows the success rate of the spread bets and the ratio of money earned/ money bet. As seen above, the success rate of spreads is 0.61, meaning that every bet we've placed as a 61 percent chance of being correct. While this success rate seems much higher than the expected (the expected rate is 50 percent), it also needs to be proven to be statistically significant. To prove that, we will perform another basic t test.

For the spreads; x = 0.609, $\mu = .5$, the $SD = \sqrt{\frac{\sigma^2}{n-1}}$, where σ^2 is the variance of a Bernoulli variable with p = .5, and n = 89.

Using these numbers, the calculated t value is 4.113, while this t value is less than the t value of the OV/UN, these findings are still statistically significant as t is still greater than 2.

3.6 Team Momentum

Many teams don't play at the same level consistently. Some teams will start off the season well, and as the season progresses they'll have more losses because teams adjust to their offensive style. Conversely, some teams will start off the season poorly, but become playoff caliber teams as the season goes on. The current formula does not account for this, as it takes into consideration the full season. To account for these trends, the formula should only cover a user defined, fixed number of previous weekly results. Once settled upon, these trailing match results should provide enough data to get accurate numbers and trends, but not enough to see the teams evolution. This evolution would also explain why the formula success rate was much higher in the first few weeks when compared to the later weeks.

After further testing, we've decided that the formula would cover a period of 5 weeks to calculate the odds results for the week, and the momentum adjustments can only be made after week 8. We've decided to delay the use of the adjustments for a few weeks to let teams adjust to their new offensive and defensive schemes. The first few weeks of the season are always filled with unexpected results and high performance variance, therefore the data is unreliable.

To test effectiveness of this strategy, we analysed 30 games in weeks 12 and 13, using our old spread and OV/UN and compared it to the adjusted spread and OVUN.

winner		loser			both off -	OV/UN (W	MOMENTUM OVUN	MOMENTUM ATS	OFF ATS
Buff	@	Pats	24	10	1	45.5 U	1	1	1
Vikes		jets	27	22	0	41 U			0
NG		Comms	20	20	1	42 U	1		1
Browns		Texans	27	14	0	44 O	1	1	1
Lions		Jags	40	14	1	49.5 O	1	1	0
Eagles		Titans	35	10	1	44 O		1	0
GB		Bears	28	19	1	45 O	1		0
Steelers		Falcons	19	16	1	36 U		1	0
Bengals		Chiefs	27	24	1	51 O	1		1
Raiders		Chargers	27	20	1	48.5 O	1	1	1
Ravens		Broncos	10	9	0	47.5 U			0
Hawks		LAR	27	24	0	42.5 U	1		0
Niners		Dolphins	33	17	1	43O	1	1	1
Boys		Colts	54	19	0	46.5 U		1	1
TB		Saints			1	U		1	0
					10		9	9	7
					450		240	240	-180
					1.272727		1.145454545	1.145454545	0.890909091

Above is the results from week 13.

The adjusted predicted odds have led to an increase in the correct amount of spread predictions and had a negligible difference in the OV/UN predictions. More testing is needed, but the preliminary results are promising.

3.7 Coding

To optimize the process further, we've created a basic code to calculate the predicted spread and score result.

````
library(readxl)
data<-read_xlsx("C:\\Users\\nadav\\Downloads\\Momentum test 13,8.xlsx")

if(nrow(data>40)){
  subset<-data[76:107,]
}else{
  subset<-data
}

teamfav <- subset[subset$Tm=="TeamA",]
teamunderdog <- subset[subset$Tm=="TeamB",]


FavOVUN<-as.numeric(teamfav$`OFF R OVUN`)
OVUNOFF<-as.numeric(teamfav$`OFF R OVUN`)+as.numeric(teamunderdog$`OFF R OVUN`)

OVUNDEF<-as.numeric(teamfav$`DEF R OVUN`)+as.numeric(teamunderdog$`DEF R OVUN`)

SpFAV<-as.numeric(teamfav$`RR+def+off*SOS`)
SpUND<-as.numeric(teamunderdog$`RR+def+off*SOS`)

OVUNind<-OVUNOFF-OVUNDEF


Spreadpred<- (SpFAV-SpUND)*13.5

OVUNline <- xx

OVUN<-function(OVUNind,OVUNline){
  if(OVUNind*yy.yy+xx>OVUNline){
    return("OVER")
  }
  else{
    return("UNDER")
  }
}

OVUN(OVUNind,OVUNline)
OVUNind*20.5+44
Spreadpred
```
````

3.8 Takeaways

In conclusion, our mathematical approach to betting utilizes an adjusted Kelly Criterion and odds ratio and is compared to the Pythagorean theorem as a benchmark. Overall, we find our method was successful in predicting the spreads and OV/UN more efficiently than the Pythagorean theorem and emotional betting. However, as a long-term betting strategy, our adjusted Kelly strategy was shown to not be profitable for money-lines. While the underdog theory worked for a multitude of weeks, the ending result was negative.

3.9 Future Research

We propose three future lines of research;

- Another addition that needs to be made is the inclusion of players missing games and the impact of their absence. In these early stages, we decided not to bet on a team that has an injured quarterback, due to the heavy negative impact of missing the star player. After further research, there
````

is a statistic known as ”win share”, and this statistic helps magnify the impact of each player and should be able to help us to reflect the strength of each team more accurately.

- Lastly, SOS isn't an accurate measure of how hard a schedules team is. To mitigate this, we would like to use an SOS that is updated weekly using the current team records.

4 Conclusions

Across both explorations, the results were mixed. Predicting NFL week-over-week had the best success, with an accuracy of around 64 percent. Then, the NFL year-over-year data performed well enough, with a test accuracy of 54.8 percent on the 2018 data and a maximum return of 26 percent. Additionally, support vector machines had the most success with machine learning tools, doing much better than regression methods or generalized additive models (GAMs). Furthermore, implementing a momentum strategy saw increased success in the weekly NFL betting strategy.

Overall, the accuracy being split as weekly NFL being the most accurate and yearly NFL least accurate seem to make sense, as there should be the least variation in a week to week setting, and that strategy performed the best. Some struggles of the models was to overcome overfitting problems. For the year-over-year strategy, using fewer predictor variables with support vector machines leads to better outcomes.

In conclusion, it is indeed possible to use mathematical, statistical, and machine learning methods to predict sports success. In the betting environment, using the Kelly Criterion leads to positive betting outcomes. Although successes with these strategies cannot guarantee future performance, the data has shown that it is possible to predict and profit better than a strategy of simply betting the underdogs or flipping a coin.

References

[1] BetIQ (teamrankings.com). Nfl over/under records in the regular season, 2014.

[2] J. Breech. 2022 nfl strength of schedule for all 32 teams: Cowboys have it easiest, rams facing toughest slate. 2022.

[3] B. James. Baseball abstract 1983. 1983.

[4] J. L. Kelly. A new interpretation of information rate. 1956.

[5] S. J. Miller. A derivation of the pythagorean won-loss formula in baseball. 2007.

[6] Pro-Football-Reference.com. 2017 nfl standings team stats, 2017.

[7] S. Nestler W. J. Winston and K. Pelechrinis. Mathletics: How gamblers, managers, and fans use mathematics in sports, second edition. 2022.

[8] Wikipedia. List of current national football league stadiums.

[9] Zojj. Example of the optimal kelly betting fraction, versus expected return of other fractional bets. 2021.

All raw data for in season optimal betting was collected from Pro-Football-Reference.com