David Kim
June 23st, 2019

I. Definition

Project Overview

This is an updated version of the NFL predictive model that I created previously. The primary objective of the model is to predict the winner of NFL games against the point spread. The point spread is essentially handicapping one team over the another to account for the perceived probability of one team winning. The goal of a point spread is to ensure that there is equal money on both teams.

After completing the first version of the NFL predictive model, I learned two important lessons. The first lesson learned is to always validate the data. The model for the first model had some data quality issues that essentially skewed the results of the model. Due to these data quality issues, the data was extracted from another source and thorough data validation was completed. The second lesson was to avoid overfitting. The first model was highly overfitted and caused the performance to vary from year to year. In order to address this, the second model used the testing accuracy of multiple years for parameter tuning to avoid overfitting and to build a generalized model.

Ultimately, this was a hobby project of mine since I have been an avid NFL fan. My goal for this model is to produce a model that consistently performs better than the baseline. I expect that the accuracy will never reach higher than 70% since the outcomes of NFL games can be highly unpredictable and the correlation between statistics vary from year to year and even week to week.

Problem Statement

This is a classification problem. The goal of the model is to predict the winner against the spread between two NFL teams. The model will use the NFL team's statistics, weather, and betting statistics to predict the winner. The target variable will predict the favorite (1) or the underdog (0). The model will be trained on several NFL seasons and tested on the current NFL season.

Metrics

For this classification problem, the accuracy will be used as the primary evaluation metric. The target variable is binary and there is no penalty on selecting one value over the other. The primary goal is to predict the correct target variable (winner). Accuracy is a value between 0 and 1 and also can be presented as a percentage. The higher the accuracy, the better.

$$Accuracy = \frac{number\ of\ items\ classified\ correctly}{all\ items\ in\ the\ dataset}$$

Data Source

The core NFL statistics data was extracted using the nfldb Python library that was created by BurntSushi that pulls NFL statistics. The library is currently unmaintained, but it works for now. The NFL data is extracted into a local PostgresSQL instance and the data is transferred to a local Microsoft SQL Server instance. Betting and weather data are manually compiled and loaded into the same Microsoft SQL Server instance. The data is aggregated and transformed using SQL to produce the final dataset. The data pipeline could use some automation, but automated sources require money, which I wanted to avoid.

NFLDB – https://github.com/BurntSushi/nfldb

NFL Betting Data - http://www.footballlocks.com/nfl_lines.shtml

NFL Weather Data - http://www.nflweather.com/

II. Analysis

Data Exploration

The initial dataset contains 190 variables. After data transformation for machine learning, the final dataset contains 253 variables. The data dictionary can be viewed using this link:

Data Dictionary

Data Analysis

The first aspect to check is the distribution of the target variable, spread_flag. The distribution is evenly distributed and not skewed. The underdog has won slightly more games than the favorite over the 2010 to 2017 seasons with a win rate of 50.8%.

Season	Favorite Won	Underdog Won	Total Games	Favorite Win %	Underdog Win %
2010	120	130	250	48.0%	52.0%
2011	118	127	245	48.2%	51.8%
2012	117	133	250	46.8%	53.2%
2013	132	116	248	53.2%	46.8%
2014	120	131	251	47.8%	52.2%
2015	112	134	246	45.5%	54.5%
2016	125	121	246	50.8%	49.2%
2017	131	114	245	53.5%	46.5%
Total	975	1006	1981	49.2%	50.8%

Over the 2010 to 2017 seasons, there are no highly correlated features variables with the target variable. The highest correlated variable is only 0.06 and the top negatively correlated variable is only -0.07. It is interesting that Green Bay and New England as a favorite are in the top 10 because Green Bay and New England have consistently had winning seasons. Also, the top correlated variable of underdog field goal percentage average over the last 5 games was unexpected.

2010 - 2017 Seasons		
Top 10 Correlated Feature Variables		
Feature Variable	**Description**	**Corr**
und_fg_percent_last5	Underdog field goal percentage average over the last 5 games	0.064896
week	Week	0.053203
fav_GB	Green Bay as a favorite	0.050986
und_score_last	Underdog score average last game	0.048196
fav_NE	New England as a favorite	0.043784
fav_fg_made_last5	Favorite field goals average over the last 5 games	0.041231
und_TEN	Tennessee as an underdog	0.040806
fav_against_und_last_ats_percent	Favorite/Underdog ATS	0.040409
und_fg_percent_last10	Underdog field goal percentage average over the last 10 games	0.040329
und_HOU	Houston as an underdog	0.040239

Top 10 Negatively Correlated Feature Variables

Feature Variable	Description	Corr
fav_to_fum_lost_last5	Favorite fumbles lost average over the last 5 games	-0.071487
fav_total_to_last5	Favorite turnover average over the last 5 games	-0.061240
fav_def_tkl_last5	Favorite tackles average over the last 5 games	-0.055688
und_NE	New England as an underdog	-0.054992
west_west	Pacific-time team playing a pacific-time road game	-0.054700
fav_total_to_last10	Favorite turnover average over the last 10 games	-0.054121
fav_to_fum_lost_last10	Favorite fumbles lost average over the last 10 games	-0.052709
fav_last_5_ats_percent	Favorite ATS record over the last 5 games	-0.043592
und_CIN	Cincinnati as an underdog	-0.039976
Fair	Fair weather	-0.038119

When reviewing the correlation matrix by year, each year has a different set of top 10 correlated feature variables. The correlation between feature variables and the target variable constantly changes from year to year. This was somewhat expected because of the unpredictability of NFL games and constant roster changes every year with new players being drafted, signed, and traded. The correlation matrices by year are available in the appendix:

Correlation Matrix by Year

For the ATS winning percentage, Green Bay has the best ATS winning percentage, 60.6%, when they are the favorite. New England is the best team against the spread when they are either the favorite or the underdog. As an underdog, New England has an ATS winning percentage of 81.3%, which is extremely high, but New England is seldom an underdog with only 16 games played as underdog out of 124 games. The ATS winning percentage data implies that it would be safe to bet on New England as a favorite and even safer as an underdog.

Top 10 Favorites by ATS %

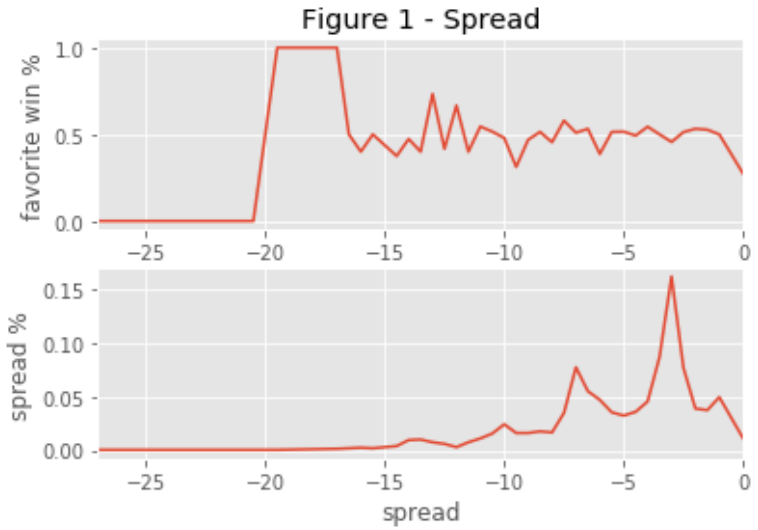
Favorite	Description	Games
GB	60.6%	94
NE	58.3%	108
HOU	56.5%	62
JAC	55.6%	27
MIN	55.4%	56
IND	54.2%	59
SEA	53.8%	78
NO	52.3%	86
CIN	52.3%	65
SF	52.3%	65

Top 10 Underdogs by ATS %

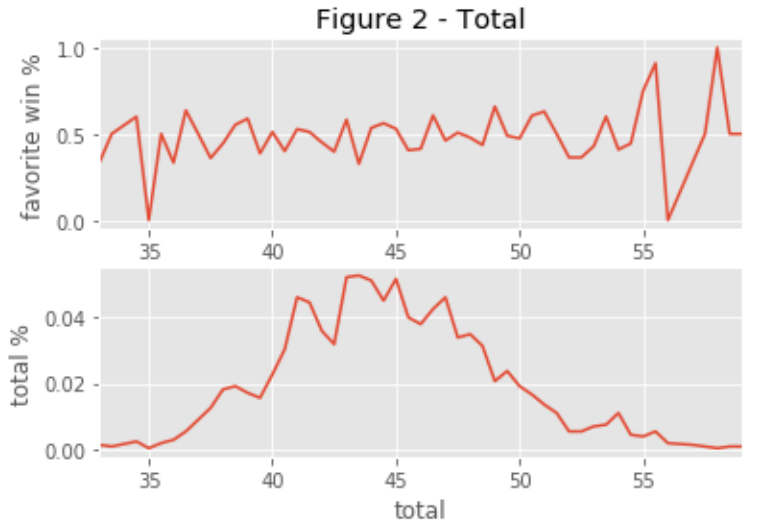
Underdog	Description	Games
NE	81.3%	16
CIN	62.5%	31
SEA	62.2%	32
MIN	59.7%	39
KC	57.9%	39
DAL	57.7%	43
NO	56.4%	45
PIT	56.3%	45
ATL	55.8%	46
SD	54.2%	52

Exploratory Visualizations

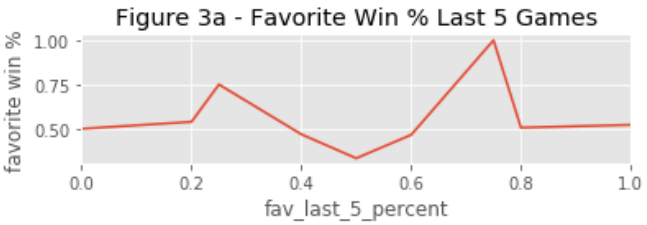
When comparing ATS winning percentage against the spread, the majority of games have a spread of 10 or less and for the most part, the ATS winning percentage hovers around 50%. There are some bigger swings between 10 to 15. For spreads between 17 to 19.5, the ATS winning percentage is 100% and any games with spreads higher than 19.5, the ATS winning percentage is 0%. Too much stock cannot be given to these observations with high spreads since there only a few games with such spreads.

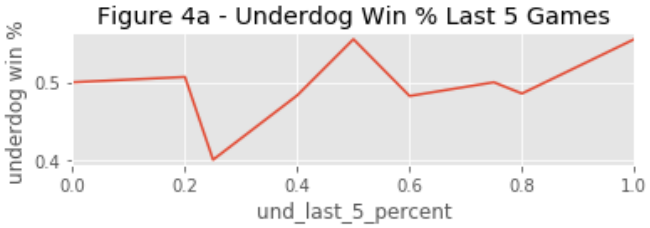


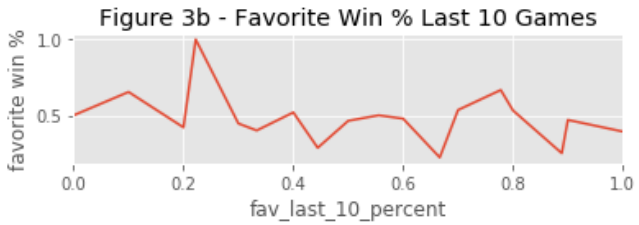
When comparing ATS winning percentage against game totals, the ATS winning percentage hovers around 50% for game totals between 36 and 54. There is a bigger variance for game totals above 55, but only 25 games are above 55. When the game total is above 55, the favorite wins at a rate of 60%.

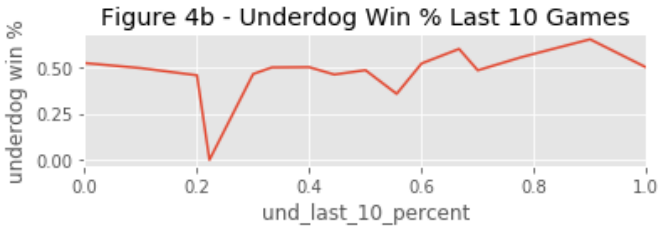


When comparing ATS winning percentage based on the favorite ATS record for the past 5 games, the data shows that favorite has a higher probability of winning when their past 5 games ATS record is between 20% and 30% or 60% and 70%. For the ATS records for the past 10 games, the data shows that the favorite has a higher probability of winning if their past 10 ATS record is between 20 and 30%.









Algorithms and Techniques

Multiple algorithms have been applied to this problem in my previous project and the XGBoost algorithm has consistently performed the best. For this project, XGBoost will be the primary algorithm. The 2010 to 2017 data will be used to train the model and the 2018 data will be considered the test data. This approach will mimic the use of the model in a real-world scenario. The number of years will be adjusted and backtested to find the optimal number of years to include in the training data. The reasoning behind this is that the older data generally will become less relevant since each year will have new rookies, player trades, and free-agent signing. Rolling 4 years, 5 years, and 6 years will be tested. For each set of years, parameter tuning will be done to find the optimal set of parameters for each year. Cross-validation will not be used since it causes the model to be highly overfitted. Instead, parameter tuning will be completed and judged based on the testing accuracy. This is a brute-force approach, but it will ensure that the model will more genialized than the previous version. Once the optimal number of years and parameters are set, feature selection will be completed to find the optimal set of features.

- Training Data
  - Rolling 4 years
  - Rolling 5 years
  - Rolling 6 years
- Parameter Tuning
  - Learning rate
  - Estimators
  - Depth
  - Subsample
  - Column sample by tree
- Feature Selection
  - Percentile

Benchmark

The benchmark model or baseline is the same as the previous version. The baseline will be selecting the underdog for all games, which will yield a winning percentage of 50.8%.

Season	Favorite Win %	Underdog Win %
2010	48.0%	52.0%
2011	48.2%	51.8%
2012	46.8%	53.2%
2013	53.2%	46.8%
2014	47.8%	52.2%

2015	45.5%	54.5%
2016	50.8%	49.2%
2017	53.5%	46.5%
Overall	49.2%	50.8%

III. Methodology

Implementation

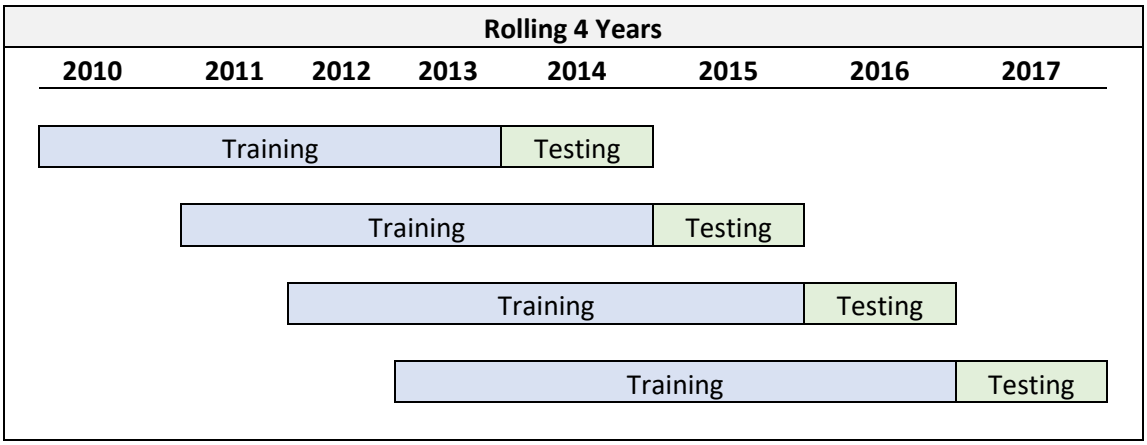
Data Preprocessing

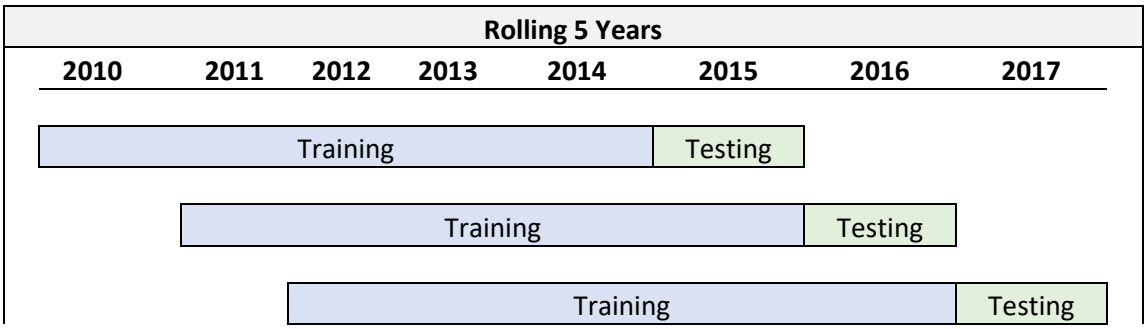
The first data preprocessing step is to remove any games that resulted in a "push", which means that the favorite team won by the spread amount, which essentially means a draw. Then the data is structured to be manually processed with the intent of use in machine learning, there are a few data preprocessing steps to account for unexpected data issues and to handle the categorical variables within the dataset. Based on the data compiled, the following steps need to take to complete data preprocessing:

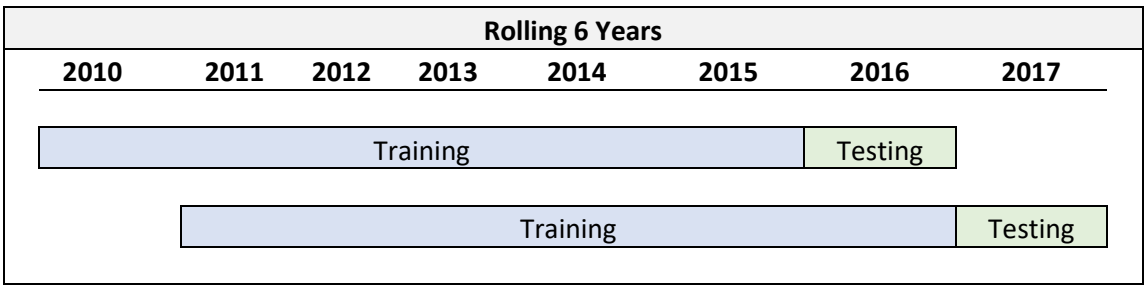
1. Fill any NaNs within the data
2. One-hot code the favorite and underdog feature variables since both variables are categorical.
3. Drop the original favorite and underdog feature variables
4. Add the one-hot coded variables into the dataset.

Implementation

The first step is to evaluate training data with a different number of years and then backtest as many iterations as possible. The different approaches for the number of years to include in the training dataset, which are a rolling 4 years, 5 years, and 6 years.







After training is completed, the model is tested against the next year. Each iteration had parameter tuning to determine the optimal set of parameters to produce the best testing accuracy and the highest minimum testing accuracy across evaluated years. The parameters tuned are the following:

- Number of trees - 10,20,30,40,50,50,60,70,80,90,100,110,120,130,140,150,160,170,180,190,200,200,300,400
- Learning rate – 0.01, 0.1,0.2,0.3
- Depth - 4,5,6,7,8,9,10
- Subsamples - 0.5,0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95,1.0
- Column samples by tree - 0.5,0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95,1.0

Once the parameters have been tuned, feature selection is completed on the model with the optimized parameters. After feature selection, the parameters and feature selection percentile are used to build a new model to test the 2018 season.

Rolling 4 Years

There were 4 sets of years completed, which were the following:

Training Set	Testing Set
2010 to 2013	2014
2011 to 2014	2014
2012 to 2015	2016
2013 to 2016	2017

The best parameters for each iteration was different and the highest testing accuracy achieved was 64.1%, which was for the set with training years of 2013 to 2016 and testing year of 2017.

Training Set	Testing Set	Trees	Learning Rate	Depth	Subsample	Col Samp by Tree	Train Acc	Test Acc
2010 to 2013	2014	130	0.2	4	0.5	0.9	**1.0000**	**0.6135**
2011 to 2014	2015	10	0.3	7	0.85	0.5	**0.9930**	**0.6138**
2012 to 2015	2016	20	0.3	10	0.95	1	**1.0000**	**0.6016**
2013 to 2016	2017	60	0.3	5	0.55	0.65	**1.0000**	**0.6408**

When comparing the parameter tuning results across each iteration, the best parameters for each set did not translate into the same results for other iterations. In order to find the best sets of parameters, the best testing accuracy across the various years. The best accuracy achieved was using the parameters for the best testing accuracy of the previous year, which was 63.7%. The next best testing

accuracy decreased as more years were added to the average: 59.2% for 2-year average, 55.8% for 3-year average, and 54.9% for 4-year average. Feature selection did not yield a different percentile other than 100%.

Years	Trees	Learning Rate	Depth	Sub sample	Col Samp by Tree	Percentile	Train 2010-2013 Test Acc 2014	Train 2011-2014 Test Acc 2015	Train 2012-2015 Test Acc 2016	Train 2012-2015 Test Acc 2017	Best Avg
4 Years	200	0.3	9	0.9	0.85	100	**0.5219**	**0.5407**	**0.5772**	**0.5551**	**0.5487**
3 Years	200	0.3	9	0.9	0.85	100		**0.5407**	**0.5772**	**0.5551**	**0.5577**
2 Years	190	0.3	5	0.55	0.65	100			**0.5650**	**0.6163**	**0.5907**
1 Year	180	0.3	10	0.9	0.6	100				**0.6367**	**0.6367**

Rolling 5 Years

There were 3 sets of years completed, which were the following:

Training Set	Testing Set
2010 to 2014	2015
2011 to 2015	2016
2012 to 2015	2017

Parameter tuning yielded a different combination of parameters for each iteration. For 2010 to 2014, the best testing accuracy was 62.2%. For 2011 to 2015, the best testing accuracy was 62.2%. For 2012 to 2016, the best testing accuracy was 63.7%.

Training Set	Testing Set	Trees	Learning Rate	Depth	Subsample	Col Samp by Tree	Train Acc	Test Acc
2010 to 2014	2015	10	0.1	7	0.6	0.6	**0.8931**	**0.6220**
2011 to 2015	2016	90	0.3	10	0.9	0.85	**1.0000**	**0.6220**
2012 to 2016	2017	60	0.3	9	0.75	0.7	**1.0000**	**0.6367**

The best accuracy achieved was using the parameters for the best testing accuracy of the previous year, which was 63.3%. Followed by the best 2-year average, which was 56.2% and best 3-year average, which was 54.7%. Feature selection did not yield a different percentile other than 100%.

Years	Trees	Learning Rate	Depth	Sub sample	Col Samp by Tree	Percentile	Train 2010-2014 Test Acc 2015	Train 2011-2015 Test Acc 2016	Train 2012-2016 Test Acc 2017	Best Avg
3 Years	170	0.3	9	0.9	0.7	100	**0.5894**	**0.5447**	**0.5061**	**0.5468**
2 Years	110	0.3	10	0.9	0.85	100		**0.6220**	**0.5020**	**0.5620**
1 Year	180	0.1	7	1	0.8	100			**0.6327**	**0.6327**

Rolling 6 Years

There were 2 sets of years completed, which were the following:

Training Set	Testing Set
2010 to 2015	2016
2011 to 2016	2017

Parameter tuning yielded a different combination of parameters for each iteration. For 2010 to 2015, the best testing accuracy was 61.0%. For 2011 to 2016, the best testing accuracy was 63.7%.

Training Set	Testing Set	Trees	Learning Rate	Depth	Subsample	Col Samp by Tree	Train Acc	Test Acc
2010 to 2015	2016	10	0.2	8	0.55	0.9	**0.9040**	**0.6098**
2011 to 2016	2017	300	0.2	9	1	0.55	**1.0000**	**0.6367**

The best accuracy achieved was using the parameters for the best testing accuracy of the previous year, which was 63.3%. Followed by the best 2-year average, which was 59.1%. Feature selection did not yield a different percentile other than 100%.

Years	Trees	Learning Rate	Depth	Sub sample	Col Samp by Tree	Percentile	Train 2010-2015 Test Acc 2016	Train 2011-2016 Test Acc 2017	Best Avg
2 Year	120	0.3	7	0.55	0.9	100	**0.6016**	**0.5796**	**0.5906**
1 Year	140	0.1	5	0.6	0.95	100		**0.6327**	**0.6327**

IV. Results

Model Evaluation

To mimic a real-world application of this model, the optimized models will be tested against the 2018 NFL season. The goal of the model evaluation against 2018 is to see how well the optimized models generalize against the next NFL season. Using the sets of years and tuned parameters, the following results are achieved:

Rolling	Yrs Inc. in Avg	Trees	Learning Rate	Depth	Sub sample	Col Samp by Tree	Percentile	Avg Test Acc 2017	Test Acc 2018
4 Years	4 Years	200	0.3	9	0.9	0.85	100	0.5487	**0.5422**
4 Years	3 Years	200	0.3	9	0.9	0.85	100	0.5577	**0.5422**
4 Years	2 Years	190	0.3	5	0.55	0.65	100	0.5907	0.5422
4 Years	1 Year	180	0.3	10	0.9	0.6	100	0.6367	0.4940
5 Years	3 Years	170	0.3	9	0.9	0.7	100	0.5468	0.5221
5 Years	2 Years	110	0.3	10	0.9	0.85	100	0.562	0.5100

5 Years	1 Year	180	0.1	7	1	0.8	100	0.6327	0.5181
6 Years	2 Year	120	0.3	7	0.55	0.9	100	0.5906	0.5301
6 Years	1 Year	140	0.1	5	0.6	0.95	100	0.6327	**0.5502**

Based on the results, the rolling 4 years using the parameters for the best average of 3 and 4 years. Even though the rolling 6 years using the parameters for the prior year produces the best testing accuracy, but it is most likely overfitted compared to the rolling 4 years models.

When comparing against the baseline, the final model's testing accuracy is 54.2%, which is 3.4% better than the baseline of 50.8%. This is a slight improvement over the baseline and is a disappointment since accuracies have reached over 60% in previous iterations. If the final model was applied in a real-world setting, the final model would yield a 7.5-unit profit at the end of the year.

$$Profit = (135\ bets\ won \times 0.9\ units\ won) + (115\ bets\ lost\ \times 1\ unit\ lost) = 7.5\ units$$

The final model's testing accuracy is actually worse than implementing the naïve baseline approach of selecting the underdog every time, which yields 55.8%. However, my expectation is the final model will perform better than the baseline and naïve approach over time.

If a model was parameter tuned and feature selected using 2018 as the testing set, the optimized model can produce testing accuracy as high as 62.3%. It would have been difficult to formulate the optimized parameters without obtaining the entire 2018 season data, which would not be possible in a real-world scenario.

Rolling	Trees	Learning Rate	Depth	Sub sample	Col Samp by Tree	Percentile	Test Acc 2018
4 Years	100	0.2	6	0.95	0.8	100	0.6225
5 Years	400	0.2	9	0.55	0.9	100	0.6145
6 Years	140	0.1	4	0.5	1	100	0.6064

V. Conclusion

Reflection

In summary, the feasible to produce a viable model for predicting NFL games against the spread. In the ideal conditions, accuracy above 60% can be achieved. However, it has been difficult to produce an approach can produce generalized models that can achieve consistent results year over year. The primary driver is that the feature variables correlation to the target variable change year over year. With every new season, new players are drafted and traded. On top of those changes, the NFL game itself is often unpredictable as well as with the numerous injuries from playing the game itself.

Improvement

There is always room for improvement since the current data set is fairly limited with data from 2010 to 2018. The most important improvement would be more data. With more data, it would be easier to backtest strategies for future years. Another improvement may be to add coach related statistics since coaches tend to have longer careers than players on average. Also, coaches appear to have a huge influence on the performance of the team. Some may argue that the coach and quarterback are the two most important people on a football team.

In an ideal situation, it would be great to find a way to capture key injuries prior to games, but there is no structured data for this data to my knowledge. On top of that, some spreads do account for these injuries if the injury information is obtained well before the game starts. Another piece of data that would be great is information on how people have bet on a particular game as well as spread (line) movements. Line movements can indicate that bettors are favoring one particular side of a game or if the information has been obtained that justify a line change. In the end, it seems that there will be always room for improvement and the model will need to be evaluated and maintained over time. The ultimate goal will be to produce a generalized model that can perform well consistently across multiple years.

Appendix

Data Dictionary

Variable Field	Variable Description	Example
gameweek	Season and year	201001
season	Season	2010
week	Week (1-17)	1
favorite_[team]	[Team] as favorite (One for each team)	1
underdog_[team]	[Team] as underdog (One for each team)	0
spread	Betting spread	-3
fav_implied_score	Favorite's implied team score based on the spread	21
und_implied_score	Underdog's implied team score based on the spread	18
total	Betting total score	39
favML	Favorite moneyline odds	-165
undML	Underdog moneyline odds	145
favoritehome	Favorite is at home	0
spreadflag	Target variable. 1 if the favorite wins and 0 if the underdog wins.	1
spread_0to3	Spread within 0 to 3 points	1
spread_35to7	Spread within 3.5 to 7 points	0
spread_75to10	Spread within 7.5 to 10 points	0
spread_105to14	Spread within 10.5 to 14 points	0
spread_145plus	Spread within 14.5 points	0
fav_as_fav_last_5_ats_percent	Favorite's ATS % as a favorite over the last 5 games	0.2
und_as_und_last_5_ats_percent	Underdog's ATS % as an underdog over the last 5 games	0.4
fav_last_5_percent	Favorite's Win-Loss % over the last 5 games	0.6
und_last_5_percent	Underdog's Win-Loss % over the last 5 games	0
fav_last_5_ats_percent	Favorite's ATS % over the last 5 games	0.4
und_last_5_ats_percent	Underdog's ATS % over the last 5 games	0.4
fav_score_last5	Favorites score average over the last 5 games	21.6
und_score_last5	Underdogs score average over the last 5 games	9
fav_spread_diff_last5	Favorites spread differential average over the last 5 games	13.4
und_spread_diff_last5	Underdogs spread differential average over the last 5 games	-6.7
fav_passing_yds_last5	Favorites passing yards average over the last 5 games	229.6
und_passing_yds_last5	Underdogs passing yards average over the last 5 games	138.8
fav_passing_yds_per_cmp_last5	Favorites passing yards per completion average over the last 5 games	10.91349
und_passing_yds_per_cmp_last5	Underdogs passing yards per completion average over the last 5 games	7.579142
fav_passing_cmp_last5	Favorites completion average over the last 5 games	0.621066
und_passing_cmp_last5	Underdogs completion average over the last 5 games	0.568736
fav_rushing_yds_last5	Favorites rushing yards average over the last 5 games	94.6
und_rushing_yds_last5	Underdogs rushing yards average over the last 5 games	97.2
fav_rushing_per_att_last5	Favorites rushing yards per attempt average over the last 5 games	4.121659
und_rushing_per_att_last5	Underdogs rushing yards per attempt average over the last 5 games	3.922344

fav_passing_tds_last5	Favorites passing TDs average over the last 5 games	1.6
und_passing_tds_last5	Underdogs passing TDs average over the last 5 games	0.6
fav_total_tds_last5	Favorites total TDs average over the last 5 games	3.2
und_total_tds_last5	Underdogs total TDs average over the last 5 games	1.2
fav_fg_made_last5	Favorites field goals made average over the last 5 games	1.2
und_fg_made_last5	Underdogs field goals made average over the last 5 games	1.6
fav_fg_percent_last5	Favorites field goal percentage average over the last 5 games	0.8
und_fg_percent_last5	Underdogs field goal percentage average over the last 5 games	0.733333
fav_xp_made_last5	Favorites extra points made average over the last 5 games	2.4
und_xp_made_last5	Underdogs extra points made average over the last 5 games	0.6
fav_def_ffum_last5	Favorite forced fumbles average over the last 5 games	0.4
und_def_ffum_last5	Underdog forced fumbles average over the last 5 games	1.4
fav_def_int_last5	Favorite interceptions average last 5 games	2
und_def_int_last5	Underdog interceptions average over the last 5 games	0
fav_def_sk_last5	Favorite sack average over the last 5 games	2.2
und_def_sk_last5	Underdog sack average over the last 5 games	1.4
fav_def_tkl_last5	Favorite tackles average over the last 5 games	54.8
und_def_tkl_last5	Underdog tackles average over the last 5 games	54.4
fav_to_fum_lost_last5	Favorite fumbles lost average over the last 5 games	1.8
und_to_fum_lost_last5	Underdog fumbles lost average over the last 5 games	0.6
fav_to_int_last5	Favorite interceptions average over the last 5 games	1.2
und_to_int_last5	Underdog interceptions average over the last 5 games	2
fav_total_to_last5	Favorite turnover average over the last 5 games	3
und_total_to_last5	Underdog turnover average over the last 5 games	2.6
fav_against_und_last_5_ats_percent	Favorite's ATS % against the underdog over the last 5 games	0.2
fav_against_und_last_5_percent	Favorite's Win-Loss % against the underdog over the last 5 games	0.4
fav_as_fav_last_10_ats_percent	Favorite's ATS % as a favorite over the last 10 games	0.3
und_as_und_last_10_ats_percent	Underdog's ATS % as an underdog over the last 10 games	0.5
fav_last_10_percent	Favorite's Win-Loss % over the last 10 games	0.6
und_last_10_percent	Underdog's Win-Loss % over the last 10 games	0.1
fav_last_10_ats_percent	Favorite's ATS % over the last 10 games	0.5
und_last_10_ats_percent	Underdog's ATS % over the last 10 games	0.5
fav_score_last10	Favorites score average over the last 10 games	23.9
und_score_last10	Underdogs score average over the last 10 games	12.1
fav_spread_diff_last10	Favorites spread differential average over the last 10 games	11.1
und_spread_diff_last10	Underdogs spread differential average over the last 10 games	-4.2
fav_passing_yds_last10	Favorites passing yards average over the last 10 games	243.1
und_passing_yds_last10	Underdogs passing yards average over the last 10 games	218.7
fav_passing_yds_per_cmp_last10	Favorites passing yards per completion average over the last 10 games	10.80424
und_passing_yds_per_cmp_last10	Underdogs passing yards per completion average over the last 10 games	10.00588
fav_passing_cmp_last10	Favorites completion average over the last 10 games	0.641694
und_passing_cmp_last10	Underdogs completion average over the last 10 games	0.587145
fav_rushing_yds_last10	Favorites rushing yards average over the last 10 games	101

und_rushing_yds_last10	Underdogs rushing yards average over the last 10 games	109
fav_rushing_per_att_last10	Favorites rushing yards per attempt average over the last 10 games	4.200115
und_rushing_per_att_last10	Underdogs rushing yards per attempt average over the last 10 games	4.187775
fav_passing_tds_last10	Favorites passing TDs average over the last 10 games	1.5
und_passing_tds_last10	Underdogs passing TDs average over the last 10 games	1.2
fav_total_tds_last10	Favorites total TDs average over the last 10 games	3
und_total_tds_last10	Underdogs total TDs average over the last 10 games	2.4
fav_fg_made_last10	Favorites field goals made average over the last 10 games	1
und_fg_made_last10	Underdogs field goals made average over the last 10 games	1.6
fav_fg_percent_last10	Favorites field goal percentage average over the last 10 games	0.8
und_fg_percent_last10	Underdogs field goal percentage average over the last 10 games	0.783333
fav_xp_made_last10	Favorites extra points made average over the last 10 games	2.3
und_xp_made_last10	Underdogs extra points made average over the last 10 games	1.2
fav_def_ffum_last10	Favorite forced fumbles average over the last 10 games	0.4
und_def_ffum_last10	Underdog forced fumbles average over the last 10 games	1.2
fav_def_int_last10	Favorite interceptions average last 10 games	1.4
und_def_int_last10	Underdog interceptions average over the last 10 games	0.2
fav_def_sk_last10	Favorite sack average over the last 10 games	2.7
und_def_sk_last10	Underdog sack average over the last 10 games	1.4
fav_def_tkl_last10	Favorite tackles average over the last 10 games	53.2
und_def_tkl_last10	Underdog tackles average over the last 10 games	51.6
fav_to_fum_lost_last10	Favorite fumbles lost average over the last 10 games	1.4
und_to_fum_lost_last10	Underdog fumbles lost average over the last 10 games	0.3
fav_to_int_last10	Favorite interceptions average over the last 10 games	1
und_to_int_last10	Underdog interceptions average over the last 10 games	1.5
fav_total_to_last10	Favorite turnover average over the last 10 games	2.4
und_total_to_last10	Underdog turnover average over the last 10 games	1.8
fav_as_fav_last_ats_percent	Favorite's ATS % against the underdog last game	0
und_as_und_last_ats_percent	Favorite's Win-Loss % against the underdog last game	0
fav_last_percent	Favorite's Win-Loss % last game	0
und_last_percent	Underdog's Win-Loss % last game	0
fav_last_ats_percent	Favorite's ATS % last game	0
und_last_ats_percent	Underdog's ATS % last game	0
fav_score_last	Favorites score average last game	7
und_score_last	Underdogs score average last game	6
fav_spread_diff_last	Favorites spread differential average last game	29
und_spread_diff_last	Underdogs spread differential average last game	-15
fav_passing_yds_last	Favorites passing yards average last game	139
und_passing_yds_last	Underdogs passing yards average last game	80
fav_passing_yds_per_cmp_last	Favorites passing yards per completion average last game	7.315789
und_passing_yds_per_cmp_last	Underdogs passing yards per completion average last game	7.272727
fav_passing_cmp_last	Favorites completion average last game	0.612903
und_passing_cmp_last	Underdogs completion average last game	0.392857

fav_rushing_yds_last	Favorites rushing yards average last game	48
und_rushing_yds_last	Underdogs rushing yards average last game	87
fav_rushing_per_att_last	Favorites rushing yards per attempt average last game	3.428571
und_rushing_per_att_last	Underdogs rushing yards per attempt average last game	3.346154
fav_passing_tds_last	Favorites passing TDs average last game	1
und_passing_tds_last	Underdogs passing TDs average last game	0
fav_total_tds_last	Favorites total TDs average last game	2
und_total_tds_last	Underdogs total TDs average last game	0
fav_fg_made_last	Favorites field goals made average last game	0
und_fg_made_last	Underdogs field goals made average last game	2
fav_fg_percent_last	Favorites field goal percentage average last game	0
und_fg_percent_last	Underdogs field goal percentage average last game	0.666667
fav_xp_made_last	Favorites extra points made average last game	1
und_xp_made_last	Underdogs extra points made average last game	0
fav_def_ffum_last	Favorite forced fumbles average last game	0
und_def_ffum_last	Underdog forced fumbles average last game	1
fav_def_int_last	Favorite interceptions average last 10 games	1
und_def_int_last	Underdog interceptions average last game	0
fav_def_sk_last	Favorite sack average last game	1
und_def_sk_last	Underdog sack average last game	3
fav_def_tkl_last	Favorite tackles average last game	56
und_def_tkl_last	Underdog tackles average last game	54
fav_to_fum_lost_last	Favorite fumbles lost average last game	0
und_to_fum_lost_last	Underdog fumbles lost average last game	0
fav_to_int_last	Favorite interceptions average last game	3
und_to_int_last	Underdog interceptions average last game	0
fav_total_to_last	Favorite turnover average last game	3
und_total_to_last	Underdog turnover average last game	0
fav_as_fav_against_und_last_ats_percent	Favorite's ATS % as a favorite against the underdog last game	1
fav_against_und_last_ats_percent	Favorite's ATS % against the underdog last game	1
fav_against_und_last_percent	Favorite's Win-Loss % against the underdog last game	1
sunday	Sunday game	1
monday	Monday game	0
thursday	Thursday game	0
saturday	Saturday game	0
morning	Morning game	0
afternoon	Afternoon game	1
night	Night game	0
central_central	Central-time team playing a central-time road game	0
central_east	Central-time team playing an east coast-time road game	0
central_mountain	Central-time team playing a mountain-time road game	0
central_west	Central-time team playing a west coast-time road game	0
east_central	East coast-time team playing a central-time road game	0

east_east	East coast-time team playing an east coast-time road game	0
east__mountain	East coast-time team playing a mountain-time road game	0
east_west	East coast-time team playing a west coast-time road game	0
mountain_central	Mountain-time team playing a central-time road game	0
mountain_east	Mountain-time team playing an east coast-time road game	0
mountain_mountain	Mountain-time team playing a mountain-time road game	0
mountain_west	Mountain-time team playing a west coast-time road game	1
west_central	West coast-time team playing a central-time road game	0
west_east	West coast-time team playing an east coast-time road game	0
west_mountain	West coast-time team playing a mountain-time road game	0
west_west	West coast-time team playing a west coast-time road game	0
Temp	Temperature at the game (Fahrenheit) Dome games set at 78	78
Dome	Game played in dome	1
Fair	Fair weather	0
Partly_Cloudy	Partly Cloudy	0
Mostly_Cloudy	Mostly Cloudy	0
Breezy	Breezy	0
Windy	Windy	0
Blizzard	Blizzard	0
Light_Rain	Light Rain	0
Heavy_Rain	Heavy Rain	0
Light_Snow	Light Snow	0
Heavy_Snow	Heavy Snow	0
Thunderstorm	Thunderstorm	0

Correlation Matrix by Year

2010 Season		
Top 10 Correlated Feature Variables		
Feature Variable	**Description**	**Corr**
mountain_west	Mountain-time team playing a pacific-time road game	0.176655
und_fg_percent_last5	Underdog field goal percentage average over the last 5 games	0.166902
und_MIN	Minnesota as an underdog	0.143751
fav_spread_diff_last	Favorite spread differential last game	0.143249
und_CAR	Carolina as an underdog	0.135588
und_last_ats_percent	Underdog ATS record last game	0.125302
und_spread_diff_last10	Underdog spread differential average over the last 10 games	0.124662
week	Week	0.121182
gameweek	Season and week	0.121182
fav_DET	Detriot as a favorite	0.114708

2011 Season

Top 10 Correlated Feature Variables

Feature Variable	Description	Corr2
fav_passing_yds_last5	Favorite passing yards average over the last 5 games	0.156506
fav_NO	New Orleans as a favorite	0.149821
fav_score_last5	Favorite score average over the last 5 games	0.148024
fav_passing_yds_last10	Favorite passing yards average over the last 10 games	0.140468
fav_def_int_last	Favorite interceptions last game	0.138075
und_LA	Los Angeles Rams as an underdog	0.136250
fav_passing_tds_last5	Favorite passing TDs average over the last 5 games	0.132992
fav_total_tds_last5	Favorite total TDs average over the last 5 games	0.132992
fav_score_last10	Favorite score average over the last 10 games	0.130615
und_fg_made_last10	Underdog field goals made average over the last 10 games	0.129313

2012 Season

Top 10 Correlated Feature Variables

Feature Variable	Description	Corr
spread_35to7	Spread is between 3.5 and 8	0.152351
fav_DEN	Denver as a favorite	0.150561
fav_rushing_yds_last	Favorite rushing yards last game	0.136100
central_mountain	Central-time team playing a mountain-time road game	0.135955
gameweek	Season and week	0.125695
week	Week	0.125695
und_BUF	Buffalo as an underdog	0.119973
und_to_int_last10	Underdog turnovers average (interceptions) over the last 10 games	0.116680
fav_IND	Indianapolis as a favorite	0.114813
Mostly_Cloudy	Mostly cloudy weather	0.111260

2013 Season

Top 10 Correlated Feature Variables

Feature Variable	Description	Corr
und_def_int_last10	Underdog interceptions average over the last 10 games	0.171665
und_passing_yds_per_cmp_last10	Underdog passing yards per completion average over the last 10 games	0.155846
und_score_last5	Underdog score average over the last 5 games	0.140858
und_score_last10	Underdog score average over the last 10 games	0.129565
und_def_int_last5	Underdog interceptions average over the last 5 games	0.127535

und_passing_yds_per_cmp_last5	Underdog passing yards per completion average over the last 5 games	0.125721
und_CHI	Chicago as an underdog	0.125413
fav_def_int_last5	Favorite interceptions average last game	0.112050
und_passing_yds_last	Underdog passing yards per game average last game	0.111411
und_passing_yds_last10	Underdog passing yards per game average over the last 10 games	0.110511

2014 Season

Top 10 Correlated Feature Variables

Feature Variable	Description	Corr
und_as_und_last_ats_percent	Underdog ATS as an underdog last game	0.179282
und_score_last	Underdog score average last game	0.143790
und_TEN	Tennessee as an underdog	0.136219
und_last_ats_percent	Underdog ATS record last game	0.127385
und_fg_percent_last5	Underdog field goal percentage average over the last 5 games	0.125706
fav_GB	Green Bay as a favorite	0.121974
und_GB	Green Bay as an underdog	0.114916
und_def_int_last	Underdog interceptions last game	0.113880
fav_def_ffum_last5	Favorite forced fumbles average over the last 5 games	0.112364
fav_IND	Indianapolis as a favorite	0.100228

2015 Season

Top 10 Correlated Feature Variables

Feature Variable	Description	Corr
und_rushing_yds_last	Underdog rushing yards last game	0.165887
und_def_tkl_last10	Underdog tackles average over the last 10 games	0.140191
und_def_tkl_last5	Underdog tackles average over the last 5 games	0.129373
fav_MIN	Minnesota as a favorite	0.126195
spread_145plus	Spread is 14.5 or higher	0.121535
fav_HOU	Houston as a favorite	0.120030
fav_CIN	Cincinnati as a favorite	0.118164
und_def_tkl_last	Underdog tackles last game	0.114241
fav_rushing_yds_last10	Favorite rushing yards over the last 10 games	0.113519
central_central	Central-time team playing a central-time road game	0.112424

2016 Season

Top 10 Correlated Feature Variables

Feature Variable	Description	Corr
fav_def_sk_last5	Favorite sack average over the last 5 games	0.183324
fav_NE	New England as a favorite	0.148770

und_to_int_last	Underdog turnovers (interceptions) last game	0.146864
fav_def_ffum_last	Favorite forced fumbles last game	0.135792
und_CLE	Cleveland as an underdog	0.127604
und_def_int_last5	Underdog interceptions average over the last 5 games	0.117590
fav_def_sk_last10	Favorite sack average over the last 10 games	0.117201
und_fg_percent_last10	Underdog field goal percentage average over the last 10 games	0.113625
fav_def_ffum_last5	Favorite forced fumbles average over the last 5 games	0.112489
und_HOU	Houston as an underdog	0.105107

2017 Season

Top 10 Correlated Feature Variables

Feature Variable	Description	Corr
und_score_last10	Underdog score average over the last 10 games	0.174721
und_spread_diff_last	Underdog spread differential last game	0.162110
und_last_percent	Underdog win/loss record last game	0.147054
und_score_last5	Underdog score average over the last 5 games	0.131044
week	Week	0.130903
gameweek	Season and week	0.130903
und_as_und_last_5_ats_percent	Underdog ATS as an underdog over the last 5 games	0.124556
fav_against_und_last_ats_percent	Favorite/Underdog ATS	0.121809
und_xp_made_last10	Underdog extra points made average over the last 10 games	0.115078
und_rushing_per_att_last	Underdog rushing yards per attempt average last game	0.111325