British Journal of Mathematics & Computer Science

22(4): 1-12, 2017; Article no.BJMCS.31565
ISSN: 2231-0851

Estimating the Strength of the Impact of Rushing Attempt in NFL Game Outcomes

Xupin Zhang[1*], Benjamin Rollins[2], Necla Gunduz[3] and Ernest Fokoué[2*]

[1] *University of Rochester, Rochester, NY 14620, USA.*
[2] *Rochester Institute of Technology, Rochester, NY 14623, USA.*
[3] *Department of Statistics, Gazi University, Ankara, Turkey.*

Authors' contributions

This work was carried out in collaboration between all authors. Author BR found and downloaded the first portion of the data. Author XZ, the corresponding author, obtained the second portion of the data, then she coordinated the thorough process of analysis. Author XZ also did the literature search, review and coordinated the write-up of the manuscript. Authors NG and EF helped with the analysis and the write-up of the manuscript. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMCS/2017/31565
Editor(s):
(1) Qiankun Song, Department of Mathematics, Chongqing Jiaotong University, China.
(2) Tian-Xiao He, Department of Mathematics and Computer Science, Illinois Wesleyan University, USA.
Reviewers:
(1) Paul Shea, Bates College, USA.
(2) Eduardo Cabral Balreira, Trinity University, USA.
(3) Marcos Grilo, Universidade Estadual de Feira de Santana, Bahia, Brasil.
(4) P. Wijekoon, University of Peradeniya, Sri Lanka.
(5) Julyan Arbel, Inria Grenoble Rhone-Alpes, France.
Complete Peer review History: http://www.sciencedomain.org/review-history/19299

Original Research Article

Received: 14th January 2017
Accepted: 2nd May 2017
Published: 2nd June 2017

Abstract

In this paper, we use estimators of variable importance from the ensemble learning technique of random forest to consistently discover and extract the knowledge that Rush Attempt is strongly related with winning football games in the NFL. Almost all researchers before us have consistently made claims of the impact/influence other statistics in the outcomes of NFL games, with Third Down Conversion Percentage and Takeaways almost universally considered as having the greatest impacts in game outcomes. Rushing as a factor of NFL success has also been mentioned, but mostly in terms of number of rushing yards per game. The novelty in this present work lies in

**Corresponding author: E-mail: xzhang72@u.rochester.edu;*

the fact that not only do we discover Rush Attempt differential to be the strongest and most dominant variable, but we also establish its dominance throughout the years, namely with 14 seasons worth of NFL games data providing firm evidence of the ubiquitous appearance of Rush Attempt at the root of every classification tree.

Keywords: NFL; rush attempt; classification trees; and random forests.

1 Introduction

The National Football League (NFL) is exceptionally popular in the United States. In 2016, American Football was still by far the most popular sport in the United States, outperforming both baseball(second) and basketball(third). There are many factors making NFL games hard to discover the factors including injuries to players, psychological factors and league rules to create parity. Numerous statisticians have been working on finding out the most important factor which determines the winner of a game. In this paper, we use machine learning techniques on 14 years worth of National Football League (NFL) data from 2000 to 2013 to discover interesting items of knowledge regarding the most dominant statistics that impact the success in NFL football games. In this study, our goal is to figure out the most important factor in deciding the outcome of a game. We use the ensemble learning methods of random forest to discover the most important features in 14 years worth of NFL data. We then create a variable importance matrix to present our results more distinctively. Finally, we explain the reasons why a single variable in football games could differentiate successful and unsuccessful teams so well.

2 Previous Research

Different machine learning techniques have been used to predict NFL games by numerous researchers. In [1] multiple neural network approaches were used to predict NFL results. His research showed that Back-Propagation (BP) was the most effective method, predicting week 16 games using 15 weeks of 1994 season with a predictive accuracy of 78.6%. More prominently, BP happens to be much better in predicting games in situations where score difference between the teams was larger.

To find the factors that distinguish dominant teams and weak ones, researchers have been using different ranking models to explain the difference in scores in NFL. A ranking model was proposed by [2] with two major assumptions: 1) whether a team wins a game is based on its interaction with its components. 2) The competitiveness of a team is determined by its ratings. Differently from the model in [2], another scholar [3] proposed a new model with an important assumption stating that a teams ratings should be in proportion to the scores they get. Several years later, [4] used an Offense-Defense approach to rank teams. Rather than having a single rating score for each team, the author aggregated two rating scores including offensive and defensive for each team. Tested on NFL, college football (NCAA football), and college basketball (NCCA basketball), the model works better than the previous models with less CPU time used in the computations. Earlier, [5] had proposed a model called Generalized Markov Model to predict the differences in scores in NFL. Compared with [4] Offense-Defense model, the major advantage of Generalized Markov Model is that it could do the calculations more efficiently by inputting several statistics at once.

Linear models have been employed by many researchers to predict NFL outcomes. As the research shows, [6] recommended linear mixed models to predict teams future performance. The good prediction was typically based on the differences in score from past games. Their average prediction error was 10.68 from $1,330$ games from 1971 to 1977. A linear model was also proposed by [7] for college football bowl prediction for 2005 using 2004 to 2006 data. They found six most important predictors making up 22% variance of the actual outcomes.

Though a majority of researchers employed quantitative methods to conduct game research, a case study was performed by [8] to discriminate successful and unsuccessful teams using FIBA 33 games. Results of the study showed that having taller players in the team might increase their possibility to win the game. Moreover, investing more time in improving aggressive offensive and defensive playing styles also helps the team become more successful.

As can be seen from the research, [9] compared rushing and passing statistics in predicting team strength. They found that passing offense outperformed rushing offense in terms of the efficiencies. Moreover, the results showed that teams with more yards per passing play have a better chance become the winner of a game. Data mining approach was employed by [10] to analyze factors that distinguish successful and unsuccessful teams in NFL. They employed data from 2009 to 2010 NFL season. By using multiple data mining and machine learning techniques including principal component analysis, factor analysis, support vector machine and logistic regression, their results showed that third down conversion is the most important factor in deciding the winner of a game in NFL.

3 Exploratory and Formal Analysis of the Data

3.1 Exploratory data analysis

The data used in this research consists of 14 years of National Football League (NFL) data from 2000 to 2013 season. The data that we used was gathered from the following site: http://www.repole.com /sun4cast/data.html. We used Microsoft Excel to organize the data. Table 1 shows the description of the variables we used in the analysis. As Table 2 shows, a typical years data included 12 variables for each home and away team. However, starting from 2010, one another variable called SackNum was added. We utilized R programming language to clean the doubles in the data. It is important to mention that the data in Table 2 consists of the differences of the statistics between the numbers of the home team and those of the visiting team. The first exploratory analysis on this data consisted in the generation of the comparative boxplots for each variable, with the goal of informally identifying which (if any) variable showed the most power (impact) in differentiating the winning team from the losing team. Fig. 1 shows such a plot for the 2012 NFL season, and reveals the emergence of difference in Rush Attempt as the clearest and most impactful separator of winners and losers.

Table. 1. Description of the variables used throughout the paper

Variable Name	Variable Description
Result	Outcome of the Game for the Home Team
First Down	Difference in the number of first downs
Third Down	Difference in third down conversion percentage
Rush Attempt	Difference in the number of rush attempts
Rush Yards	Difference in the number of rushing yards
Pass Attempt	Difference in the number of pass attempts
Pass Completion	Difference in the number of pass completions
Pass Yards	Difference in the number of passing yards
Pass Interception	Difference in the number of pass intercepted
Fumble	Difference in the number of fumbles
Sack Number	Difference in the number of sacks

Table. 2. View of the Game Statistics for 16 of the 512 games played during the 2012 NFL season

Result	First Down	Third Down	Rush Att	Rush Yds	Pass Att	Pass Comp	Pass Yds	Pass Int	Fumble	Sack Num
Win	3	7.00	7	61	-3	1	103	1	-1	-1
Win	0	-1.00	-13	-72	2	2	71	0	0	-2
Win	-1	-19.00	-10	-68	-2	2	51	-2	-1	-2
Loss	-1	-4.00	-10	77	5	-1	-71	2	1	0
Loss	-2	-16.00	-23	-120	9	7	163	2	0	1
Win	4	13.00	18	51	-10	-2	21	-2	-2	-1
Loss	-13	-27.00	-8	-51	-21	-17	-195	0	-1	0
Win	1	-2.00	1	19	-14	-3	31	-1	1	-3
Win	14	11.00	-9	5	23	15	173	3	0	-2
Loss	-1	24.00	-18	-141	18	10	88	1	0	-1
Win	4	21.00	16	4	-5	0	58	-3	-1	-1
Loss	2	30.00	5	-10	12	3	-24	0	0	0
Win	9	14.00	19	142	-12	-6	-36	-1	-1	-1
Loss	2	-9.00	-34	-121	26	5	15	2	1	1
Win	6	6.00	-5	-7	-5	1	115	-1	-1	-1
Loss	6	-5.00	0	13	13	8	50	0	1	2

Even after combining all the 14 seasons considered in this research, the difference in Rush Attempt remains by very far the most powerful separator of winners and losers. This finding, although informal at this point (since we haven't yet considered a formal statistical model), appears to confirm a popularly held perception (belief), that rushing is crucial to success in the game of football. Many actually believe that even great passing teams first establish the run to open passing lanes through various strategies and plays like play action. It is interesting therefore for consider rigorous statistical methods to further find out if indeed the difference in Rush Attempt does play such a central role in the outcome of NFL games. To achieve such a formal analysis aim, we can consider traditional techniques and methods. Since we have a binary classification situation in this case, one might be tempted to first consider logistic regression and combine it with state of the art techniques of variable selection to determine the most powerful statistics in an NFL game. In this paper however, we thought it better to use classification trees, first because of their interpretability, but also because of their Random Forest extension that provides a very good estimation of the importance of variables.

Fig. 1. provides the first informal view into the importance of each variable. Each boxplot depicts a separate variable. It can be seen that Rush Attempt is (at least informally for now) the factor that achieves the clearest classification of the NFL game outcome, separating winner from loser more clearly than any other variable.

3.2 Classification trees analysis

Classification and Regression Trees have been widely studied and applied to a wide variety of data mining problems. Essentially, given explanatory variables $\mathbf{x}_i = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{ip})^\top \in \mathcal{X} \subseteq \mathbb{R}^p$ and response variables $Y_i \in \{1, \cdots, G\}$ making up the data set $\mathcal{D} = \Big\{(\mathbf{x}_1, Y_1), \cdots, (\mathbf{x}_n, Y_n)\Big\}$, classification trees perform pattern recognition by partitioning the input space $\mathcal{X}$ with respect to $\mathcal{Y}$ into q regions $R_1, R_2, \cdots, R_q$ so that the corresponding tree is the union of all the regions, namely $\texttt{Tree} = \cup_{l=1}^{q} R_\ell$. The resulting learning machine (approximating function), is a piecewise constant function such that given a new point $\mathbf{x}^{\texttt{new}}$, its predicted response (class) is given by

$$\hat{Y}^{\texttt{new}}_{\texttt{Tree}} = \widehat{\texttt{g}}_{\texttt{Tree}}(\mathbf{x}^{\texttt{new}}) = \sum_{l=1}^{q} \left\{ I(\mathbf{x}^{\texttt{new}} \in R_l) \left\{ \underset{j \in \{1,\cdots,G\}}{\texttt{argmax}} \left\{ \frac{1}{|R_l|} \sum_{\mathbf{x}_i \in R_l} I(Y_i = j) \right\} \right\} \right\}. \tag{1}$$

As indicated earlier, we first considered all the 14 NFL seasons separately and built the corresponding classification tree. We then combined all the seasons into one single dataset and built the classification tree for the combined data.

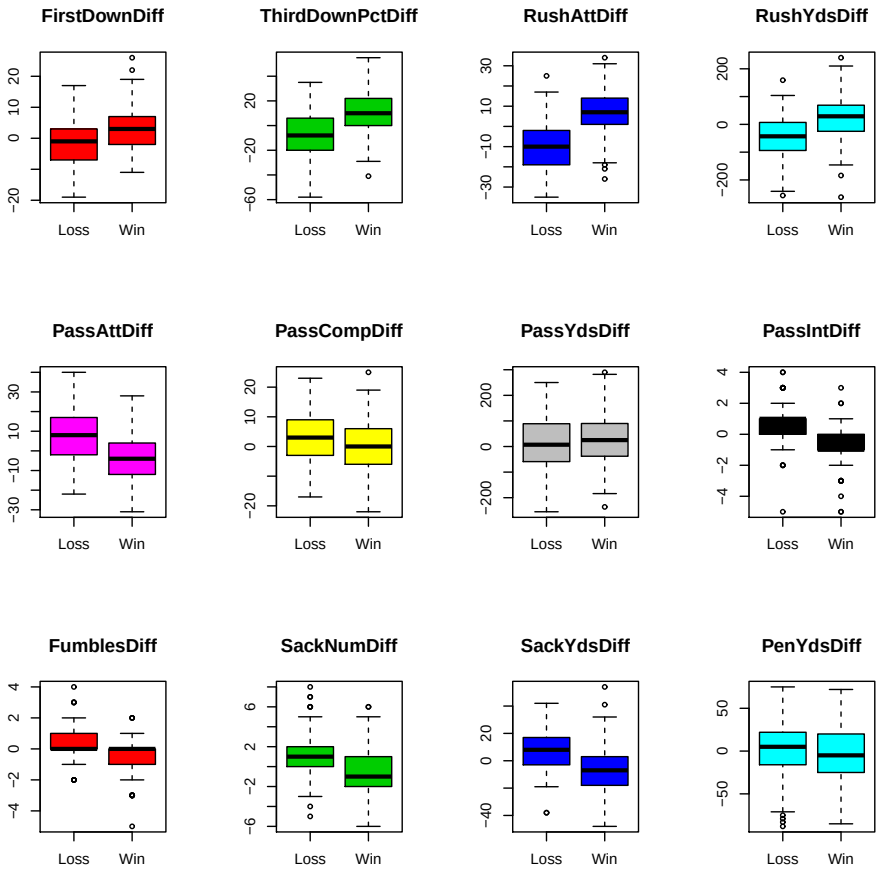
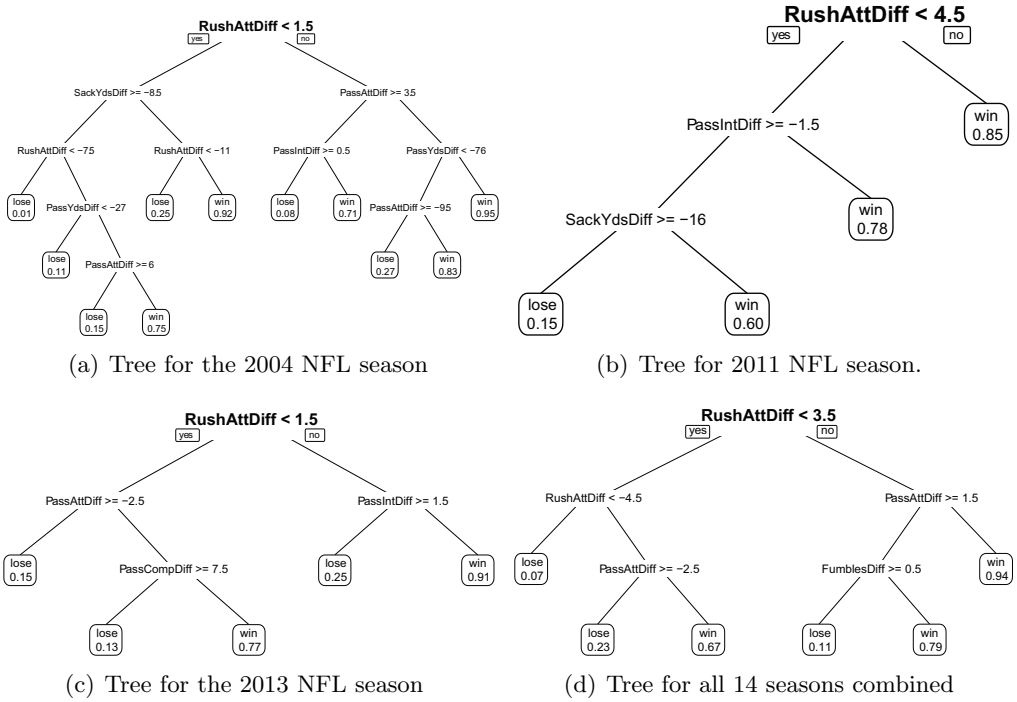


Fig. 1. Comparative Boxplots for each of the 12 variables on which game statistics were recorded

The pattern depicted in Fig. 2 reveals the absolute dominance of `Rush Attempt` as the most impactful factor in the outcome of NFL games. We actually generated classification trees for all the 14 seasons analyzed in this paper, and for all those seasons, the variable Rush Attempt Difference was the root of the tree. We only showed 3 of the years in the interest of space. It appears clear from both the individual seasons and the combined seasons that the trees declare Rush Attempt the most impactful variable in determining the outcome of NFL games. Considering the fact that trees are notorious for being unstable (high variance), we thought that its determination of Rush Attempt as the superior variable might be questioned.

Fortunately, the random forest (RF) classifier which is a natural extension of the tree classifier, provides a built-in mechanism for estimating the importance of variables. Besides, since a random forest is an ensemble of trees and thereby a more stable learning machine, we decided to analyze our data with it, and confirm or (maybe infirm) the previously noticed dominance of Rush Attempt.



(a) Tree for the 2004 NFL season

(b) Tree for 2011 NFL season.

(c) Tree for the 2013 NFL season

(d) Tree for all 14 seasons combined

Fig. 2. Classification trees for several NFL seasons

3.3 Random forest analysis

A statistical machine learning forest [11], just like a real life forest, is an ensemble of trees. [12] first introduced the algorithms for random decision trees. [13] developed random forests by using out-of-bag error as an estimate of the generalization error and measuring variable importance through permutation. In recent years, the ensemble learning technique of random forest have been an important topic in the field of biostatistics, engineering and many other fields [14]. For our work, we used the R package randomForest [15] to perform our analyses and predictions for NFL data. Recall that we have class labels y coming from $\mathscr{Y} = \{1, 2, \cdots, G\}$ and predictor variables $\mathbf{x} = (\mathtt{x}_1, \mathtt{x}_2, \cdots, \mathtt{x}_p)^\top$ coming from a p-dimensional space $\mathscr{X}$. Let $\widehat{\mathtt{g}}^{(\mathtt{b})}(\cdot)$ be the bth bootstrap replication of the estimated base classifier $\widehat{\mathtt{g}}(\cdot)$, such that $(\widehat{\mathtt{y}})^{(b)} = \widehat{\mathtt{g}}^{(\mathtt{b})}(\mathbf{x})$ is the bth bootstrap estimated class of $\mathbf{x}$. The estimated response by Random Forest (RF) is obtained using the majority vote rule, which means the most frequent label throughout the B bootstrap replications. Succinctly, we can write the RF estimated label of $\mathbf{x}$ as

$$\widehat{f}^{(\mathtt{RF})}(\mathbf{x}) = \underset{\mathtt{y}\in\mathscr{Y}}{\mathtt{arg\,max}} \left\{ \mathtt{freq}_{\widehat{\mathtt{C}}^{(\mathtt{B})}(\mathbf{x})}(\mathtt{y}) \right\} = \underset{\mathtt{y}\in\mathscr{Y}}{\mathtt{arg\,max}} \left\{ \sum_{b=1}^{B} \left(\mathbf{1}_{\{\mathtt{y}=\widehat{\mathtt{g}}^{(\mathtt{b})}(\mathbf{x})\}} \right) \right\}. \tag{2}$$

In greater details,

$$\widehat{f}^{(\mathtt{RF})}(\mathbf{x}) = \mathtt{Most\ frequent\ label\ in\ } \widehat{\mathtt{C}}^{(\mathtt{B})}(\mathbf{x}), \tag{3}$$

where

$$\widehat{\mathtt{C}}^{(\mathtt{B})}(\mathbf{x}) = \left\{ \widehat{\mathtt{g}}^{(1)}(\mathbf{x}), \widehat{\mathtt{g}}^{(2)}(\mathbf{x}), \cdots, \widehat{\mathtt{g}}^{(\mathtt{B})}(\mathbf{x}) \right\}. \tag{4}$$

Algorithmic details of Random Forest are well known and vastly documented and can be found in [13]. Table 3 depicts the Random Forest Variable Importance plots for 6 of the 14 NFL seasons considered. Consistently and quite uniformly, Rush Attempt emerges as the most important variable, and even appears to be consistently far ahead of any other variable coming second to it.

It can be seen from Figs 3 and 4 that Rush Attempt emerges as the most important of all the variables, which should not surprise here, since it was the top in each of the seasons.

Table. 3. Rank of random forest variable importance

	Rush Att	Pass Att	Third Down	Sack Yds	Pass Int	First Down	Fumbles	Pen Yds	Pass Yds	Rush Yds	Pass Comp
2000	1	6	5	2	3	7	10	9	8	4	11
2001	1	3	11	4	5	6	7	8	9	2	10
2002	1	6	5	2	3	7	10	9	8	4	11
2003	1	6	5	2	3	7	10	9	8	4	11
2004	1	2	5	2	3	5	9	8	6	4	11
2005	1	3	10	4	6	5	11	9	8	2	7
2006	1	5	10	3	2	6	7	8	9	4	11
2007	1	5	10	3	2	6	7	8	9	4	11
2008	1	5	10	3	2	6	7	8	9	4	11
2009	1	3	9	5	2	7	11	8	6	4	10
2010	1	2	10	5	3	6	11	7	8	4	9
2011	1	4	8	3	2	7	11	9	6	5	10
2012	1	4	6	2	3	7	10	9	8	5	11
2013	1	3	10	5	2	6	11	7	8	4	9
Combined	1	3	9	5	2	7	10	8	6	4	10

3.4 Logistic regression

In the interest of completeness, we thought it useful to explore another learning machine different from classification trees and random forest. To that end, we fitted a linear logistic regression model to the data and performed the bidirectional stepwise procedure for model selection. We specifically used the traditional logistic regression formulation that assumes that the response variable Y_i is related to the explanatory vector $\mathbf{x}_i$ through the model

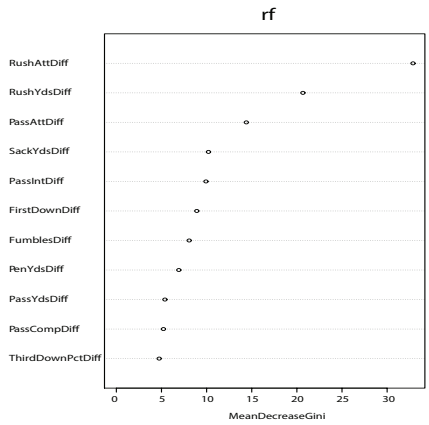
$$\log\left[\frac{\pi_i}{1-\pi_i}\right] = \eta(\mathbf{x}_i;\boldsymbol{\beta}) \tag{5}$$

where $\eta(\mathbf{x}_i;\boldsymbol{\beta}) = \mathbf{x}_i^\top\boldsymbol{\beta} = \beta_0 + \beta_1\mathbf{x}_{i1} + \beta_2\mathbf{x}_{i2} + \cdots + \beta_p\mathbf{x}_{ip}$ and

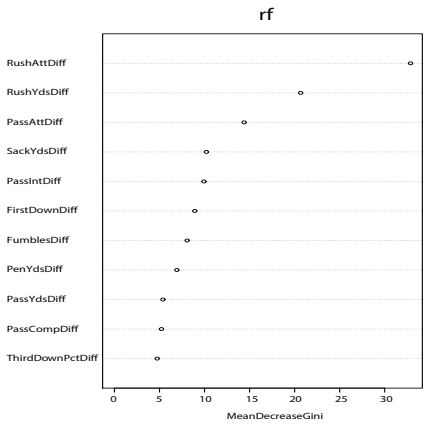
$$\pi_i = \Pr[Y_i = 1|\mathbf{x}_i] = \frac{e^{\eta(\mathbf{x}_i;\boldsymbol{\beta})}}{1+e^{\eta(\mathbf{x}_i;\boldsymbol{\beta})}} = \frac{1}{1+e^{-\eta(\mathbf{x}_i;\boldsymbol{\beta})}} = \pi(\mathbf{x}_i;\boldsymbol{\beta}) \tag{6}$$

Below is the out from R.

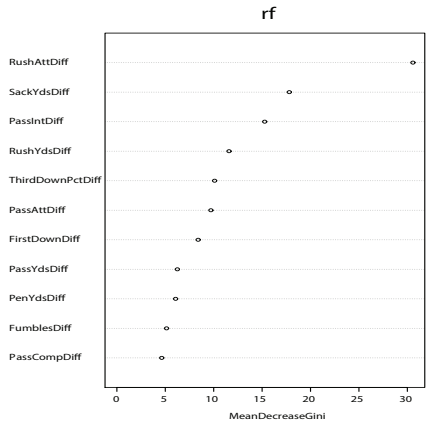
```
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.237213   0.228341  -1.039 0.298873
FirstDownDiff     0.204891   0.050074   4.092 4.28e-05 ***
ThirdDownPctDiff  0.044233   0.014299   3.093 0.001979 **
PassAttDiff      -0.199615   0.034758  -5.743 9.31e-09 ***
PassYdsDiff       0.016293   0.004253   3.831 0.000127 ***
PassIntDiff      -0.920326   0.209554  -4.392 1.12e-05 ***
FumblesDiff      -1.176667   0.280184  -4.200 2.67e-05 ***
PenYdsDiff       -0.026114   0.007237  -3.609 0.000308 ***
```



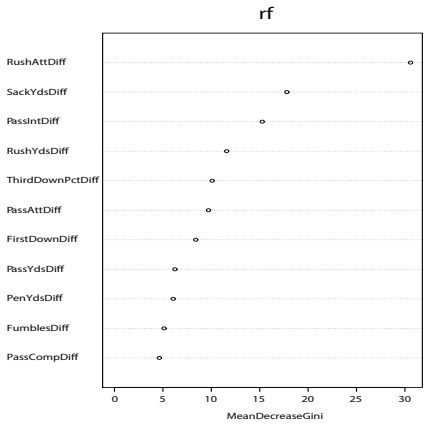
(e) 2000 NFL season.



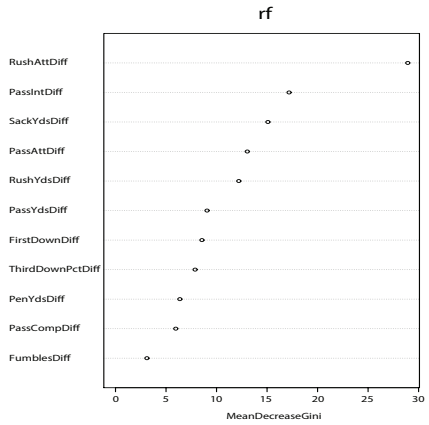
(f) 2001 NFL season



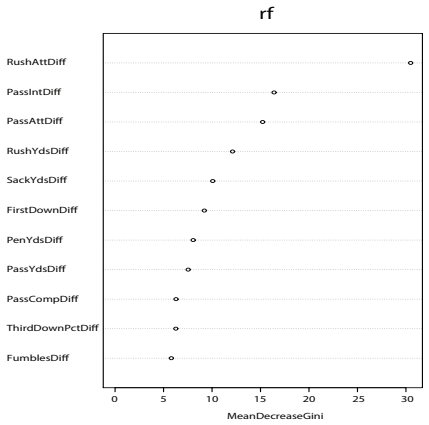
(g) 2002 NFL season.



(h) 2004 NFL season



(i) 2011 NFL season.



(j) 2013 NFL season

Fig. 3. Random Forest Variable Importance Plots for Several of the 14 NFL seasons analyzed

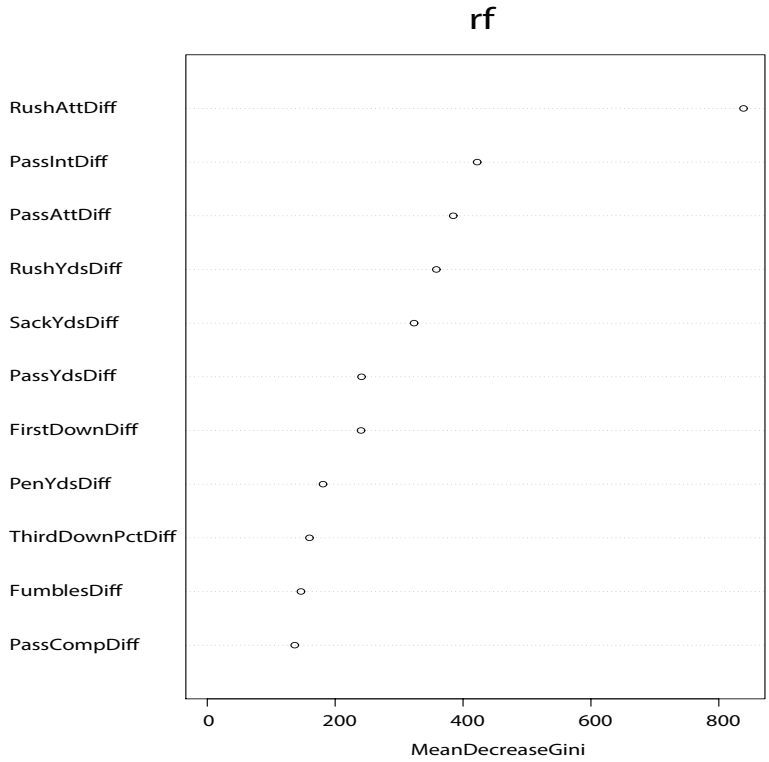


Fig. 4. Random Forest Variable Importance Plot for the combined data of the 14 NFL Seasons Analyzed

It is quite noteworthy that Rush Attempt which uniformly dominated the tree and random forest analyses is not even present in the model. Upon deep analysis of the reasons, it becomes clear that this absence is not surprising. The extreme importance of Rush Attempt as revealed by Random Forest, caused it to be an **absorbing** explanatory variable, a characteristic that does not impede coordinate-based learning machines like trees, but seriously affects global methods like generalized linear model. Essentially, what happens when a variable is overwhelmingly strong absorbing, is that it tries to be the only predictor of the response at the exclusion of the rest. When that happens a techniques like stepwise will tend not retain such a variable. Now, in the context of our NFL data, the absence of Rush Attempt in the GLM is not alarming, precisely because, of the remaining variables left in the final model, the top variable Pass Attempt is the one that came third and the one right after it Pass Interception is the one that came second, in the Random Forest Variable Importance estimation of the data made of all the seasons combined. When one examines the GLM variables carefully, the findings related to Rush Attempt are further confirmed. Precisely, The Z-value for Pass Attempt is -5.743 (largest in magnitude of all the variables), and the Z-value of Pass Int is -4.392, which are both very large in the negative direction. Clearly, this agrees with the power of Rush Attempt established earlier, in the sense that winning teams pass less and are less intercepted than losing teams. This was even clearly apparent in Fig. 1. Despite this dominance of rush attempt as established by random forest, we noticed a weird phenomenon with GLM, Rush Att was entirely out. Via both GLM and traditional stepwise, our explanation is that Rush Att Diff

is dominant alone that combine with other variables. Fig. 5 shows the predictive performance of Random Forest, Linear Discriminant Analysis, and Quadratic Discriminant Analysis.

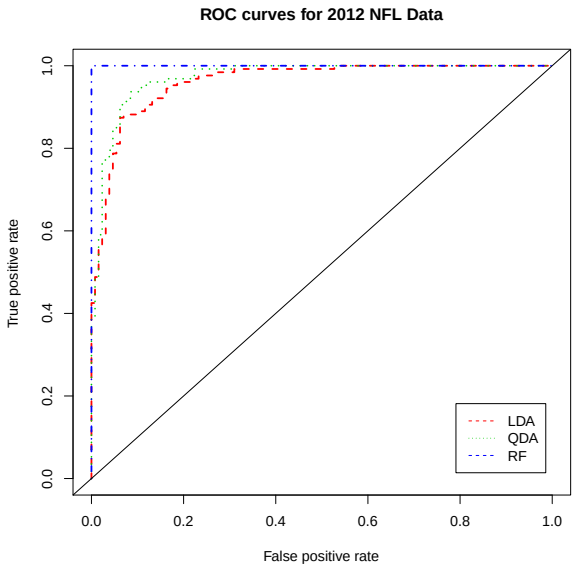


Fig. 5. ROC Curve for Classification of 2012 NFL Data

4 Discussion and Conclusion

As illustrated throughout the explanations given above via both single tree classifiers and random forest classifiers, we established that Rush Attempt outperforms all the other 10 variables in each year from 2000 to 2013. From our findings, a natural question arises as to how Rush Attempt would perform in prediction if considered alone.

Random Forest provided us with estimates of variable importance. In our case, the dominance of Rush Attempt was established, somewhat providing a window into the tactical dimension of NFL gamesmanship. Indeed, logistic regression on 14 years worth of data, appears to suggest that in a randomly chosen single NFL game, an increase of 1 Rush Attempt difference over an opponent increases to the odds of beating that opponent by 17%. One could tactically advise a coach to work harder on developing a team around good rushing to the point of maximizing well designed rush attempts in games to increase chances of victory.

This paper never intended to claim to have definitively established a cause-effect relationship between rush attempt and the outcome of NFL games. Rather, we sought and successfully demonstrated with strong and compelling evidence from 14 seasons worth of NFL data, that Attempt is clearly uniformly strongly associated with the outcome of any NFL game. In fact, from a predictive perspective, the ROC curve of Fig. 5 is yet another piece of evidence to support our conclusion.

From Figs. 3 and 4, it can be seen that the dominance of Rush Attempt differential is obvious and complete. The variable importance matrix of Table 3 is also consistent with our previous findings,

namely that Rush Attempt outperforms all the other variables throughout the 14 seasons analyzed in this paper. Besides, the pattern seen here in 14 years individual ones repeats itself in the merged data set, which confirms our assumption that Rush Attempt is the most important factor that seems to characterize the dominant/winning teams.

Many authors have done quantitative analyses of NFL data, and both those authors and NFL managers, NFL pundits, NFL reporters would argue about the importance of rushing, but never before has anyone spotted the dominance of rush attempt. Those who ever spoke about rushing associated with NFL game outcome, concentrate on Rushing Yards. Our work has demonstrated with crystal clarity that it is Rush Attempt and not Rushing Yards that seems to make the bigger (in this case the hugest) difference. In that sense, this work is very novel and highlight for the first time the importance of an NFL game statistic not deem crucial until now. In his 2012 article [16] entitled Passing League: Explaining the NFL's aerial evolution, author Steve Wyche is attempting to demonstrate that the superiority of NFL teams, and therefore the important factor in their success (game outcome) is strongly associated with the passing prowess of quarterbacks connecting with elite receivers. Many other authors and pundits and casual NFL enthusiasts, siding with [16], have declared the death of the rushing dominance, even lamenting the misfortune of teams predicated on the old philosophy of 3 yards and a cloud of dust championed by the legendary Ohio State university Head Coach Woody Hayes. Our work appears to indicate that rushing is as stronger as ever, and that not even rushing yards but surprising the mere Rush Attempt holds some of the keys to the success in the great game of American Football. Although we have not established a cause-effect relationship between Rush Attempt differential and NFL game outcome, the uniform consistency of our empirical demonstration of the dominance of Rush Attempt warrants deeper studies into its predictive impact on NFL games.

As illustrated throughout the explanations given above via both single tree classifiers and random forest classifiers, we established that Rush Attempt outperforms all the other 10 variables in each year from 2000 to 2013. From our findings, a natural question arises as to how Rush Attempt would perform in prediction if considered alone.

Random Forest provided us with estimates of variable importance. In our case, the dominance of Rush Attempt was established, somewhat providing a window into the tactical dimension of NFL gamesmanship. Indeed, logistic regression on 14 years worth of data, appears to suggest that in a randomly chosen single NFL game, an increase of 1 Rush Attempt difference over an opponent increases to the odds of beating that opponent by 17%. One could tactically advise a coach to work harder on developing a team around good rushing to the point of maximizing well designed rush attempts in games to increase chances of victory.

Competing Interest

Authors have declared that no competing interests exist.

References

[1] M. Purucker. Neural network quarterbacking. IEEE Potentials. 1996;15:9-15.

[2] James P. Keener. The perron-frobenius theorem and the ranking of football teams. SIAM Review. 1993;35:90-93.

[3] Massey K. Statistical models applied to the rating of sports teams; 1997.

[4] Carl D. Meyer, Amy N. Langville, Anjela Y. Govan. Offense-defense approach to ranking team sports. Journal of Quantitative Analysis in Sports. 2009;5.

[5] Anjela Y. Govan. Ranking theory with application to popular sport. PhD thesis, North Carolina State University; 2008.

[6] David Harville. Predictions for national football league games via linear-model methodology. Journal of the American Statistical Association. 1980;75(371):516-524.

[7] Brady West, Lamsal Madhur. A new application of linear modeling in the prediction of college football bowl outcomes and the development of team ratings. Journal of Quantitative Analysis in Sports. 2008;4(3).

[8] Koh KT, John W, Mallett C. Discriminating factors between successful and unsuccessful teams: A case study in elite youth olympic basketball games. Journal of Quantitative Analysis in Sports. 2011;7(3).

[9] May C, Dan M, Carl M, Ralph A, John H. Rush versus pass: Modeling the nfl. Journal of Quantitative Analysis in Sports; 2010.

[10] Fokoue E, Foehrenbach D. A statistical data mining approach to determining the factors that distinguish championship caliber teams in the national football league; 2013.

[11] Andy Lia, Matthew Wiener. Classification and regression by randomforest. R News. 2002;2(3):18-22.

[12] Ho TK. Random decision forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition. 1995;278-282.

[13] Breiman L. Random forests. Machine Learning. 2001; 45(1):5-32.

[14] Gromping U. Variable importance assessment in regression: Linear regression versus random forest. The American Statistician. 1996;63(4):308-319.

[15] Andy Liaw, Matthew Wiener. Classification and regression by randomforest. R News. 2002;2(3):18-22.

[16] Steve Wyche. Passing league: Explainning the NFL's aerial evolution; 2012 (accessed July 5, 2012).
Available: http://www.nfl.com/news/story/09000d5d82a44e69/article/passing-league-explaining-the-nfls-aerial-evolution

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/19299