Model for Maximizing Punt Return Yardage
NFL Big Data Bowl 22

Andrew Curtis
a02043584@aggies.usu.edu
Utah State University

ABSTRACT

For the NFL Big Data Bowl 2022, the challenge was to take data from special teams plays and develop new strategies to implement. Using provided metrics on players and play operation, several predictors for kick return yardage were put to the test. Kick length, returner height and hang time emerged as the best prospects for predicting kick return yardage. Using a machine learning linear regression approach, a reduced model was developed and is presented. Longer kicks tend to favor the kicking team, but receiving teams should still return these as there is a larger potential for return as well. Shorter returners are favored for the receiving team. Perhaps of just as much interest to teams are the factors that did not have significant effects on return yardage. Although there is a high amount of variance in kick returns, it was possible to develop a significant predictive model and use it to deploy new special teams strategies.

CCS CONCEPTS

• **Machine Learning** → **Linear Regression**.

KEYWORDS

datasets, linear regression, machine learning

1 INTRODUCTION

The NFL Big Data Bowl 22, hosted by Kaggle, features data centered on special teams plays for the 2018-2020 seasons. A common special teams play type is kicks to opposing teams. In these plays, offensive possession of the ball is traded from one team to the other, by having a kicker advance the ball downfield. The receiving team will then attempt to receive the ball and advance it back for as many yards as possible, this begins their offensive possession of the ball. There are two main types of kick plays that get returned, kickoffs and punts. In this model, only punts are considered for reliability of predictions. However, the opportunity is similar between these two types of plays and the player a team chooses to return the ball is often the same for both types.

An initial investigation of the data set lead to a question of what type of player physique would be most suited to maximizing returns. Kick returner height and weight could be matched to return yardage. There are a few different conventional archetypes for the type of player physiques that are most suited to maximize return yardage. One might be that a larger, stronger runner can power through multiple defenders and fight for more yards. However, a more modern consensus is that a small, light and fast runner will tend to be more evasive and can gain more yards.

In addition, more predictors were later added that were primarily dependent on the kicking team. These were kick length, hang time, operation time and snap time. Using all of the play related metrics available, a more complete model was investigated. With both sides of the kick represented, possible strategies can be developed for both teams.

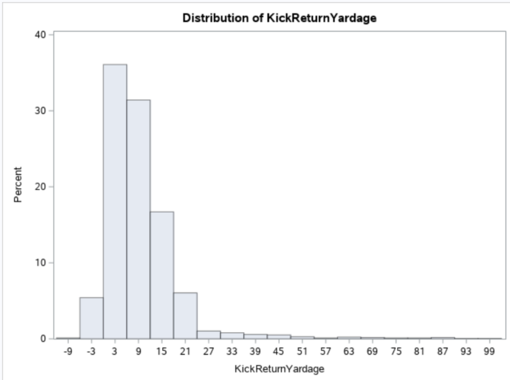
2 DATA EXPLORATION AND CLEANING

This project utilized the "players" and "plays" data sets from the NFL Big Data Bowl 2022. The players data set contains information on height, weight, age, NFL Id, and several other variables. The NFL Id is a player-specific Id and was used to map players between data sets. In the plays data set, special teams plays are tracked with many different variables to describe the individual plays. For this project, kick return yardage was chosen as the response variable.

For pre-processing, kicking plays, specifically punts were selected. Fair catches were removed. Fair catches are where the kick returner chooses not to attempt to advance the ball. Additionally, the punts were checked with the kick direction intended and actual variables to confirm that the punt was not mishit unexpectedly. The purpose of these limitations were to remove unwanted outliers and influential points that would affect a model. The kick return yardage response variable should not include these types of plays, as it is not indicative of a returner's ability to gain yards. The players and plays data sets were joined using the NFL Id and the Returner Id, respectively. Additionally, height was converted from various forms to integer values of inches.

Below are smoothed response plots for kick length, hang time, returner height and returner weight. In general, increasing kick length is positively correlated with kick return yardage. Hang time and returner height appear to be negatively correlated. Notably, returner weight does not have as obvious of a relation. Some of these results may seem counter-intuitive at first, but represent and interesting insights into new analysis for kick returns.

The regression model presented considers 1,754 punts. The mean distance for a kick return was 8.6 yards and had a standard deviation of 10.3 yards. The worst return was -8 yards and the maximum was 99 yards. A histogram of kick return yardages is plotted below,

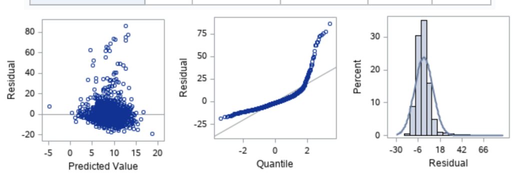


3 DEVELOPING THE LINEAR MODEL

The predictor variables were regressed on kick return yardage and an ANOVA test was run. The results of the ANOVA test, as well as residual error plots and numerical tests for normality and constant variance of residuals are shown below.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	8887.37619	1481.22936	14.56	<.0001
Error	1747	177693	101.71348		
Corrected Total	1753	186581			



P-value for Brown-Forsythe test of constant variance in residual vs. predicted

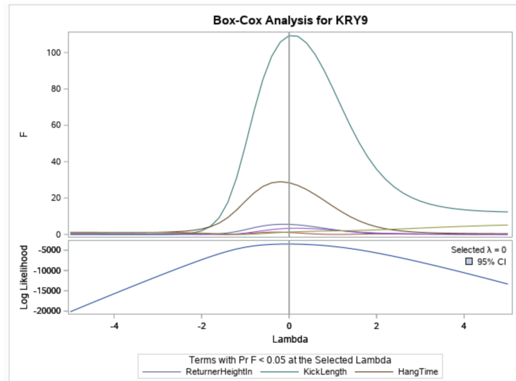
Obs	t_BF	BF_pvalue
1	3.53190	.000423265

Pearson Correlation Coefficients, N = 1754
Prob > |r| under H0: Rho=0

	resid	expectNorm
resid residual	1.00000	0.85583 <.0001
expectNorm	0.85583 <.0001	1.00000

The results of the ANOVA test show that the model is significant. However, the residuals appear to possibly have non-constant variance with some outliers. From the Residual vs. Quantile plot, the residual errors veer significantly from the expected indicating possible issues with normality. The histogram plot may also indicate some violation of the normality of residuals. The Brown-Forsythe test shows that the constant variance of residuals is violated. The Pearson Correlation Coefficient test for normality of residuals shows that the assumption of normality of residuals is also violated. Based on these results, we consider a power transformation.

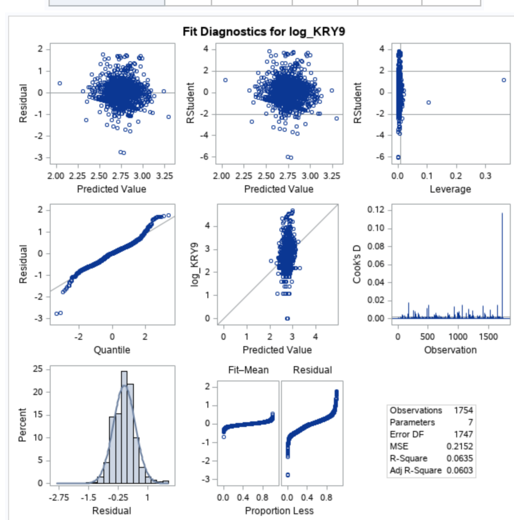
A Box-Cox analysis was run and is presented below. From this analysis a log power transformation was chosen to be applied. Note, the kick return yardage values were adjusted by +9 yards on all return data points in order to correctly run the Box-Cox analysis and apply the transformation. This was in order to fix returns with negative yardage gained.



After the transformation, the residual vs. predicted value plot shows significant improvement for outliers, constant variance and normality. The quantile and histogram plots also show significant improvements. Parameter estimates for the full model are also included.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	25.49229	4.24872	19.75	<.0001
Error	1747	375.89778	0.21517		
Corrected Total	1753	401.39007			



Parameter Estimates							
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Type I SS**	**Squared Partial Corr Type I**
Intercept	1	3.61369	0.47088	7.67	<.0001	13248	.
ReturnerHeightIn	1	-0.01690	0.00717	-2.36	0.0185	1.32467	0.00330
ReturnerWeightLb	1	0.00116	0.00112	1.04	0.3004	0.14297	0.00035737
KickLength	1	0.01879	0.00180	10.46	<.0001	17.42662	0.04357
HangTime	1	-0.14508	0.02720	-5.33	<.0001	5.83782	0.01526
OperationTime	1	-0.18573	0.10134	-1.83	0.0670	0.47750	0.00127
SnapTime	1	0.22302	0.19456	1.15	0.2518	0.28271	0.00075153

The numerical tests for the transformed model show that constant variance and normality of residuals are upheld. Interaction and higher order terms were tested and were determined to be insignificant. There were also no multicollinearity issues present based on variance inflation and condition index numerical tests.

For variable selection metrics, the adjusted r-squared variable analysis was used. The top results for this analysis are shown below.

Number in Model	Adjusted R-Square	R-Square	Variables in Model
6	0.0603	0.0635	ReturnerHeightIn ReturnerWeightLb KickLength HangTime OperationTime SnapTime
5	0.0603	0.0629	ReturnerHeightIn KickLength HangTime OperationTime SnapTime
5	0.0601	0.0628	ReturnerHeightIn ReturnerWeightLb KickLength HangTime OperationTime
4	0.0600	0.0622	ReturnerHeightIn KickLength HangTime OperationTime
4	0.0595	0.0616	ReturnerHeightIn ReturnerWeightLb KickLength HangTime
3	0.0594	0.0610	ReturnerHeightIn KickLength HangTime
5	0.0590	0.0617	ReturnerHeightIn ReturnerWeightLb KickLength HangTime SnapTime
4	0.0589	0.0611	ReturnerHeightIn KickLength HangTime SnapTime
4	0.0582	0.0603	KickLength HangTime OperationTime SnapTime

From the adjusted R-square analysis, we choose the top 3-variable model. This includes Returner height, kick length and hang time. Notice that this model is not the top adjusted R-square score, but for a small trade-off in captured variance the 3-variable model can be chosen. This choice greatly increases interpretability.

4 FINAL MODEL

A final reduced model is presented below. All predictors are significant and the reduced model captures a large amount of the full model variance. Notice, this is still with the log transformation of kick return yardage. The model assumptions are also upheld.

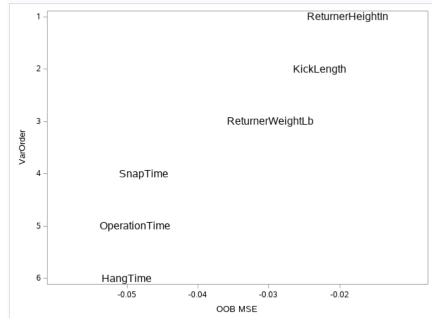
$$\log(\text{KickReturnYardage}) = \beta_o + \beta_{1,Height}\text{X}_1 + \beta_{2,KickLength}\text{X}_2 + \beta_{3,HangTime}\text{X}_3 + \epsilon$$

$$\log(\text{KickReturnYardage}) = 3.3105 - 0.0125\text{X}_1 + 0.0186\text{X}_2 - 0.1407\text{X}_3$$

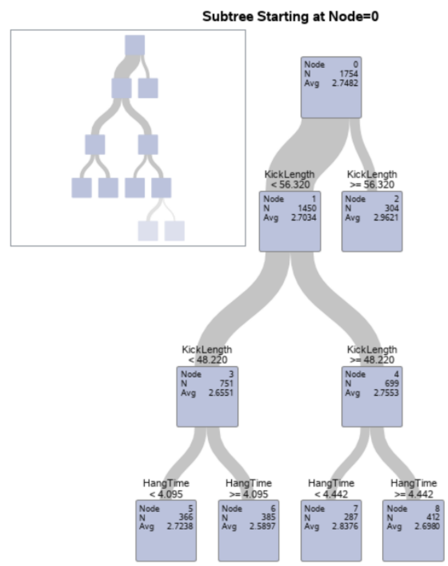
To test the predictive capability of the reduced model, the transformed data was split into a train and test set. The mean square error of the train set was 0.215. The value for the MSPR of the test set was 0.230. This measure being only slightly higher indicates that the predictive error is low and the reduced model has good performance.

5 ALTERNATIVE METHODS

Two alternative methods of regression tree and random forest were applied to the data. The results of the random forest are shown below.



The parameters of returner height, kick length and returner weight are the 3 most important variables to the regression. Interestingly, these are two of the 3 variables selected for the final model from the multiple linear regression. The results of the regression tree are shown below.



The regression tree has two splits on kick length and two splits of hang time. The result of these alternative methods is strong evidence that kick length is the main predictor in terms of importance.

6 RESULTS

The final reduced model with included variables of returner height, kick length and hang time captured most of the variance contained by the full model. The following sections provide realistic examples of the effect on kick return yardage that a team could expect.

A six-inch increase in returner height is associated with a decrease of 1.08 kick return yards. Six inches is a fairly large disparity for teams to cover. Perhaps the maximum difference between player heights on a team may be about 16 inches. However, choosing a shorter returner represents a consistent increase to return yardage for teams.

An increase of 5 yards of kick length equates to a gain of 1.1 yards on the return. This example can be partially explained by conventional special teams wisdom, but confirmation and fitting numerical values would be quite important to teams. Increasing kick length starts the receiving team in a worse position but correlates to increased return yardage. As the trade-off is about 5-to-1, the kicking team would much rather have a longer kick. This also informs the receiving team that if it is a longer kick, they should probably return it. As opposed to conducting a fair catch, since are more expected yards to be gained.

One second of increased hang time is associated with a decrease of 1.15 kick return yards. A possible explanation for this result is that if the kicking team has more time to close in on the returner; they can better limit return yardage. However, the regression tree provides an interesting analysis of hang time and kick length. It appears that there may be a "sweet spot" of hang time that produces the best kick return yardage. Potentially, kick returners could count this off in their head to know if a return is going to fall in this region.

The amount of variance in punt plays is high, with many players on the field interacting not to mention extraneous conditions, like weather and competitive pressure. These models do not capture exceptionally large amounts of the variance but are able to derive interesting results and predict potentially consistent and useful improvements to special teams performance.

7 CONCLUSION

Conventional NFL special teams wisdom states that a small, fast runner will produce better results as a kick returner. It is also commonly thought that a long high kick will result in the least amount of return yards. However, from fitting a variety of predictors predictors for a linear regression model to kick return yardage, some of these old adages may be challenged.

Although there is a high amount of total variance with returns a reduced model was able to fit kick length, returner height and hang time as significant predictors in the final selected model. From the train/test analysis, the reduced model is shown to have good predictive capability for the data. Increased kick length is associated with an increase in kick return yardage. This could perhaps be explained with conventional knowledge that a longer kick can give more separation to a runner. However, our model shows that this increase in kick return yardage is mitigated, as the yards gained from the kick outweigh the increased return by about 5 to 1. Hang time is negatively correlated with return yardage. A possible explanation is that the kicking team has more time to close in on the runner to stop them. However, the results of the regression tree shows that there may be a sweet spot for hang time that maximizes return yardage. This information could possibly be used by the runner to gauge whether a fair-catch should be called. The final predictor, returner height, has a negative correlation with return yardage. This is similar to conventional wisdom, but interestingly weight does not play much of a factor. Perhaps the smaller height of the returner has some hidden features in influencing the plays. These results challenge some conventional special teams wisdom and represent new features to adjust to develop strategies for the benefit of both the kicking and receiving teams on kickoff plays.