EXTRA POINT UNDER REVIEW

Machine Learning and the NFL Field Goal

JAMES LEDOUX

THE NEW NFL EXTRA POINT RULE FIRST IMPLEMENTED IN THE 2015 SEASON REQUIRES A KICKER TO ATTEMPT HIS EXTRA POINT WITH THE BALL SNAPPED FROM THE 15-YARD LINE. THIS ATTEMPT STRETCHES AN EXTRA POINT TO THE EQUIVALENT OF A 32-YARD FIELD GOAL ATTEMPT, 13 YARDS LONGER THAN UNDER THE PREVIOUS RULE. THOUGH A 32-YARD ATTEMPT IS STILL A CHIP SHOT TO ANY PROFESSIONAL KICKER, MANY NFL ANALYSTS WERE SURPRISED TO SEE THE NUMBER OF EXTRA POINTS THAT WERE MISSED. SHOULD THIS REALLY HAVE BEEN A SURPRISE, THOUGH? BEGINNING WITH A REPLICATION OF A STUDY BY CLARK ET. AL, THIS STUDY AIMS TO EXPLORE THE WORLD OF NFL KICKING FROM A STATISTICAL PERSPECTIVE, APPLYING ECONOMETRIC AND MACHINE LEARNING MODELS TO DISPLAY A DEEPER PERSPECTIVE ON WHAT EXACTLY MAKES SOME FIELD GOAL ATTEMPTS MORE DIFFICULT THAN OTHERS. ULTIMATELY, THE GOAL IS TO GO BEYOND THE PREVIOUS RESEARCH ON THIS TOPIC, PROVIDING AN IMPROVED PREDICTIVE MODEL OF FIELD GOAL SUCCESS AND A BETTER METRIC FOR EVALUATING PLACEKICKER ABILITY.

DATA SELECTION

The data used in this experiment comes from Armchair Analysis, covering every field goal attempted from the 2000 through 2013 seasons, roughly 14,000 attempts in total. It contains the binary outcomes of field goal attempts, atmospheric conditions at kickoff time (temperature, precipitation, humidity, wind speed), and various situational variables, such as kick distance, time left in the game, the game's score, and whether the kicker was iced. While there is no way to predict probability with perfect precision, the size of this data set allows certain models to form predictions that are reasonably close to the true value on this 0-to-1 scale of make or miss.

Here is an overview of the most significant information in the data set:

Distance:

Distance is the most important variable in this data set. The distances in the data range from 18 to 76 yards, with the inner quartile range being from 25 to 45 yards, and the median distance being 37 yards. The distribution of distances in the data set is more or less uniform from 20 to 50 yards, but sparse beyond 55.

Stadium Factors:

The data also includes the stadium in which the kick took place, whether it took place on turf or natural grass, and whether this happened at a home or an away game. Most of these situational factors proved insignificant, and specific stadiums were not included in the final models. There was, however, significant proof to support the much-discussed "Mile-High Effect," under which kickers experience significantly increased range capabilities while playing in the high altitude of the Denver Broncos' Mile High Stadium.

Climate Factors:

Among the climate factors measured were temperature, humidity, wind speed, and precipitation. The distributions of these factors are presented below:

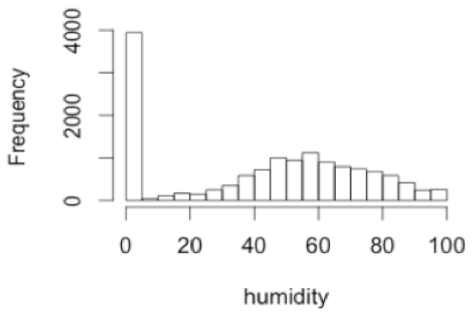

FIGURE 1. TEMPERATURE DISTRIBUTIONS IN THE DATA

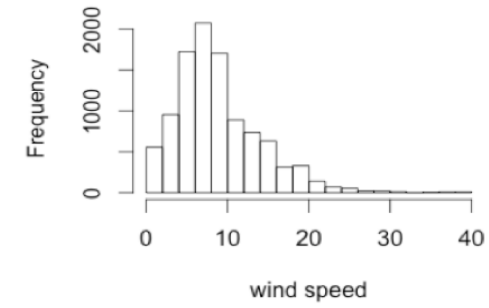

FIGURE 2. WIND DISTRIBUTIONS IN THE DATA

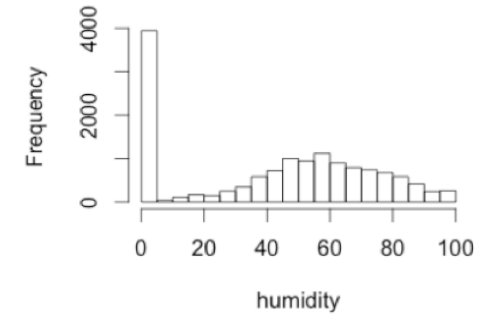

FIGURE 3. HUMIDITY DISTRIBUTIONS IN THE DATA

Given these distributions, I carried out similar feature engineering to that done in the Clark paper, creating a binary "cold" dummy variable for temperatures under 50°F, a "windy" variable for wind speeds greater than 10 mph, and a "humid" one for humidities greater than 60%. The cold and windy dummy variables are useful because both wind speeds and temperatures vary throughout games, while the measurements in this data set are only those recorded

at kickoff time. Converting these measurements to categorical variables limits the number of cases that are misclassified due to this limitation in the data set. It should be noted that I do, however, use continuous temperature as a variable in certain later models, with temperatures typically being stable within a reasonable range to make the continuous variable useful despite its time-of-measurement flaw. Humidity was not significant in any of the models, and was dropped.

Also noteworthy is that the data set contained a variety of classifications for weather conditions such as "rain", "snow", "light rain", "stormy", "foggy", "clear", and so on. To reduce the ambiguity, I lumped all rain and snow-containing conditions into a single dummy variable called "precipitation." Of this, the set has 847 instances with precipitation, and 13,144 without.

There were many other variables to test in the data set, but these proved to be the most important to the end-models. The individual effects of each of these variables on the final predictive models can be seen in the appendix, exhibit B.

BINOMIAL LOGISTIC REGRESSION MODELS

The first model I tested was a logistic regression model, replicating that used by Clark et al., and experimenting with some slight adjustments.

Logistic regression is a non-linear translation of the standard MLR function, measuring probability between the only two possible field goal outcomes of a make and a miss (1 and 0). This probability is calculated by the function displayed at the bottom of this page.

The Clark paper removes all regressors not statistically significant at the 5% level (P value < .05). This leads it to only include distance and a series of environmental factors, omitting all situational ones such as point spread, the kicker being iced, the home field, and etc. I tested three models, the first being a perfect replication of the Clark model, the second including the icing of the kicker, and the third including statistically significant home-field advantages.

Variable	Coefficient	Std. Error	P-Value
Intercept	5.5232	0.1215	0.0000
Distance (yards)	-0.1024	0.0027	0.0000
Cold (<50°F)	-0.2430	0.0690	0.0004
Precipitation	-0.3205	0.0944	0.0006
Windy (≥10 mph)	-0.2040	0.0509	0.0000
Turf	0.2809	0.0487	0.0000
Altitude (≥4000 ft.)	0.6068	0.2709	0.0250

TABLE 1. CLARK REPLICATION REGRESSION OUTPUTS

Model 1: Clark Replication

My first task was to replicate the Clark study with more recent data. Clark's model uses data from 2000-2011, so I replicated this with the 2012 and 2013 seasons included to see if there were any significant changes. The coefficients were all close to those from the original study, and each of the variables either became more significant or remained the same. The outputs are shown in Table 1 and can be compared to the originals in exhibit C.

Model 2: Icing the Kicker

After replicating the study, I included the variable of icing the kicker. This is a binary variable equal to one when the opposing team had called a timeout just before the field goal was attempted in a potentially game winning or tying situation, and equal to zero in all other cases. This was excluded in the Clark model because it did not meet the significance level of 5%. Although this is still the case in my model, I argue that a 9.1% significance level is sufficient

$$P(make = 1) = \frac{1}{1+e^{-(\beta_0+\beta_{dist}X_1+\beta_{cold}X_2+\beta_{precip}X_3+\beta_{windy}X_4+\beta_{turf}X_5+\beta_{altitude}X_6+\beta_{iced}X_7)}}$$

"Field goal kicking, however, being the isolated component of play that it is, has the potential to be one of the first great leaps in the quantification of the ways that coaches, scouts and fans view the game."

for a binary variable such as this one, given the importance of icing kickers to the outcomes of games. This model is discussed in greater detail below.

Model 3: Home Field Advantages

For model 3, I regressed all of the previously mentioned variables as well as the individual stadium dummies on field goal outcomes in order to see which home fields were the most significant. I then iteratively removed the least significant stadiums until I arrived at a list that met the 10% significance level, relaxing my significance levels slightly in order to get a better view of home field advantages in the data set.

The stadiums that were significant at the 10% level and their coefficients are: Adelphia Coliseum (-.3567), Giants Stadium (-0.5444), Network Associates Coliseum (-0.6063), Paul Brown Stadium (-0.2335), Ralph Wilson Stadium (-0.3736), Texas Stadium (-0.3096), and Veterans Stadium (-0.1536).

The reason most stadiums were insignificant appears to be the high correlation between many stadiums and their associated weather conditions. Lambeau Field, Soldier Field, and Gillette Stadium, for example, were all storied stadiums with reportedly difficult crowds to play against. These stadiums also, however, have high levels of precipitation and historically cold temperatures. This problem with collinearity led me to discard this model altogether, with the intuition that each stadium dummy is going to be highly correlated with its home-city's climate conditions and the type of surface of its playing field.

Of the three logistic regression models tested, I found the second model, with icing the kicker included, to be the most useful. Though it does not quite meet the same standard of significance as the Clark model, the importance of icing the kicker and the explanatory power it offers in game-winning and tying scenarios–one of the most enticing times to use such a model–makes this a worthwhile sacrifice. Additionally, one sub-10% significance level is certainly not problematic in a model such as this one, where everything else meets a high standard of proof.

With this established as my model of choice, the regression output is as follows:

Variable	Coefficient	Std. Error	P-Value
Intercept	5.5089	0.1361	0.0000
Distance (yards)	-0.1026	0.0030	0.0000
Cold (<50°F)	-0.2170	0.0780	0.0053
Precipitation	-0.3264	0.1062	0.0021
Windy (≥ 10mph)	-0.1730	0.0573	0.0025
Turf	0.3301	0.0548	0.0000
Altitude (≥ 4000ft)	0.8225	0.3213	0.0104
Iced	-0.1730	0.1024	0.0913

TABLE 2. MODEL 2 REGRESSION OUTPUT

When tested against data it had not seen before, here is the layout of the model's predictions measured against their distances:

Figure 4 shows that the presence of the various environmental factors and the icing of the kicker creates a considerable spread in field goal difficulties, even within the same distance-levels. A 50-yard field goal, for example, can

have anywhere between a 41 and 72 percent probability of success given this wide array of non-distance factors.

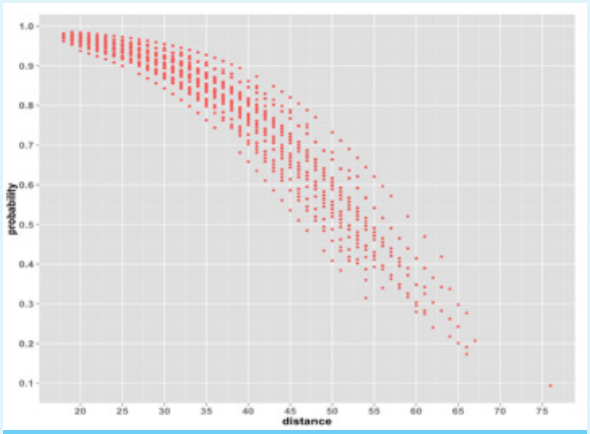

FIGURE 4: LOGISTIC REGRESSION MODEL PREDICTIONS (PROBABILITY THAT A FIELD GOAL IS SUCCESSFUL). FOR A LARGER VIEW, SEE EXHIBIT D

CONSIDERING FIELD GOAL DIFFICULTY WHEN RATING KICKER ABILITY

The most common metric used to evaluate kicker performance is "make percentage." While this is an acceptable surface-level measurement, it makes the false assumption that all kicks are created equal, which is simply not the case. Based on team strategy and quality, some kickers may attempt a greater portion of their field goals from longer distances than others. Also, due to climate factors, a kicker in New England will typically have a more difficult job than one in Oakland, all other variables held equal. Because of factors such as these, make percentage is a lacking metric; a statistic taking field goal difficulty into account would tell a deeper story.

Such a metric can be formed using the logistic regression model created above. To do this, one simply takes the difference between each field goal's outcome (0 for miss, 1 for make) and its predicted probability of success, as determined by the model, and multiplies this by 3, the number of points that the field goal is worth if successful. This creates a "points added" metric, showing the number of points that the kicker has generated for his team above the expected value, or, in other words, the value that the kicker has created above that which would be generated by a perfectly average replacement. This standardized-points measure punishes kickers greater amounts for missing "easy" field goals than it does difficult ones, overcoming the shortcomings of using the make percentage statistic alone.

Points Added = (field goal success – probability of success) x 3. This metric, measuring field goal outcomes against their difficulty level, will provide an unbiased assessment of a kicker's skill.

Who are the best kickers of the modern era?

Using this statistic, one can determine not only who have been the best kickers of the modern era, but also the most underrated, taking the difference between kickers' rankings according to the points-added statistic and those determined by make percent.

Kicker	Points-Added Per Attempt	Pct. Made	Points-Added Rank	Percent Rank (FG%)	Degree Underrated	Attempts (2000-2013)
Justin Tucker	0.372615842	0.923	1	1	0	78
Dan Bailey	0.304864407	0.908	2	2	0	98
Kai Forbath	0.30107944	0.875	3	4	1	40
Blair Walsh	0.282920419	0.898	4	3	-1	69
Greg Zuerlein	0.198175934	0.830	5	24	19	59
Rob Bironas	0.19545974	0.855	6	10	4	283
Connor Barth	0.187260008	0.841	7	19	12	120
Robbie Gould	0.166262695	0.863	8	7	-1	278
Steven Hauschka	0.15511815	0.872	9	5	-4	125
Dan Carpenter	0.145006194	0.838	10	21	11	192

TABLE 3: BEST KICKERS BY POINTS ADDED, FG%, AND DEGREE UNDERRATED

Sebastian Janikowski and Stephen Gostkowski are two of the great stories of this metric.

Gostkowski is widely touted as the best kicker in the game today. Much of his praise is well earned: he is consistently among the most accurate kickers in the NFL as measured by field goal percent (in the two most recent seasons not in this data set, he scored on 94.6 and 91.7% of his attempts). However, Gostkowski also plays for the high-powered New England offense, and part of his high make percentage appears to result from him being set up for large amounts of easier field goals. Because of this, while he is ranked 8th in field goal percentage, he is only 23rd in points-added, making him overrated by 15 positions on the list. Granted, given his successful 2014 and 2015 seasons, it is likely that these ranks have improved since 2013.

Kicker	Points-Added Per Attempt	Pct. Made	Points-Added Rank	Percent Rank	Degree Underrated	Attempts
John Carney	-0.031961265	0.829861111	42	25	-17	288
Garrett Hartley	-0.046479034	0.825688073	46	29	-17	109
Mike Vanderjagt	0.032939242	0.850961538	28	12	-16	208
Stephen Gostkowski	0.047780284	0.859848485	23	8	-15	264
Gary Anderson	0.013872664	0.844262295	31	16	-15	122
Morten Andersen	-0.021497815	0.82967033	41	26	-15	182
Matt Bryant	0.034029428	0.848387097	27	14	-13	310
Matt Stover	0.069996328	0.868656716	17	6	-11	335
Mike Nugent	-0.061526922	0.8125	48	37	-11	208
Lawrence Tynes	-0.092412104	0.805555556	53	42	-11	252

TABLE 4: WHO IS MOST OVERRATED? (RANKED HIGHER BY FG% THAN BY POINTS ADDED)

Kicker	Points-Added Per Attempt	Pct. Made	Points-Added Rank	Percent Rank	Degree Underrated	Attempts
Sebastian Janikowski	0.094237196	0.801339286	14	44	30	448
Randy Bullock	-0.021045149	0.742857143	40	63	23	35
Caleb Sturgis	0.01259131	0.764705882	32	54	22	34
Josh Scobee	0.064687163	0.808118081	19	40	21	271
Greg Zuerlein	0.198175934	0.830508475	5	24	19	59
Connor Barth	0.187260008	0.841666667	7	19	12	120
Dan Carpenter	0.145006194	0.838541667	10	21	11	192
John Kasay	0.122759211	0.838509317	11	22	11	322
Ryan Succop	0.043994456	0.813333333	25	36	11	150
Mason Crosby	0.001348368	0.794354839	36	47	11	248

TABLE 5. WHO IS THE MOST UNDERRATED? (RANKED HIGHER BY POINTS ADDED THAN BY FG%)

Janikowski, on the other hand, plays for the woeful Oakland Raiders, who have struggled offensively throughout the past several seasons and relied greatly on Janikowski's powerful leg during the period tested. As a result, he was set up with a disproportionately high amount of long field goal attempts, and his 80% field goal accuracy from 2000-2013 (ranked 44th) does not tell the full story of his impact. Due to his ability to convert from 50+ yards, his points-added score is quite high at .09/attempt, making him underrated by 30 positions on the list.

Who had the best seasons?

Finally, one can also use this statistic to measure which kickers have had the greatest impacts on individual seasons. Taking the sum of a kicker's points added by year gives his total points added for each season, showing exactly how valuable a kicker was to his team each year. Here are the greatest single-season performances according to this measurement:

Kicker	Season	Team	Points Added	Season Attempts
Neil Rackers	2005	ARI	19.684	42
Sebastian Janikowski	2009	OAK	18.926	29
Sebastian Janikowski	2011	OAK	17.371	35
Mike Vanderjagt	2003	IND	16.595	40
Steven Hauschka	2013	SEA	15.781	43
Justin Tucker	2013	BAL	15.726	41
Stephen Gostkowski	2013	NE	15.591	42
Rob Bironas	2011	TEN	15.340	32
Phil Dawson	2013	SF	15.161	42
Blair Walsh	2012	MIN	14.574	39

TABLE 6: WHICH KICKERS HAVE THE GREATEST IMPACTS ON INDIVIDIDUAL SEASONS

All of the above kickers were crucial to their teams' offenses these seasons. As you can see, one reason many people appear on this list is high kick quantity, which is to be expected – kickers who are relied upon more heavily have increased chances of accumulating points-added for their teams. This makes it all the more impressive that Sebastian Janikowski's 2009 season makes number two on this list with only 29 attempts. Although the Raiders were a losing team this season, they would have been far worse off had it not been for Janikowski bailing out their poor offense from beyond the 50 yard line multiple times over, netting an additional 19 points that the Raiders would not have secured with a replacement kicker.

IMPROVING PREDICTIONS WITH MACHINE LEARNING

Finally, while there is a great deal to be learned from the logistic regression model and the measurements that follow from its predictions, I could not resist the temptation of taking this a step further to see what additional insight machine learning models could provide in this situation. Using the same data, I tested random forest, neural network, multiple linear regression, and hybrid models to see how they fared against the logistic regression benchmark.

Random Forests

A "random forest" is an ensemble-learning algorithm that uses a resampling method called "bagging" to take a single data set and break it into several overlapping sets. With one data set now broken into many, the algorithm can train an entire "forest" of decision trees rather than just one, and uses the output of each tree as a "vote" toward what the forest's output should be. By taking the average of the many "decision trees" outputs, the random forest model achieves a lower model variance than an individual decision tree while maintaining a similar level of bias, meaning that it is both more stable and equally accurate. In this case, however, the random forest underperformed the other models tested, and did not prove useful for prediction improvement. This may be due to a combination of unreliable measurements in the explanatory variables and the fact that the majority of explanatory variables were categorical rather than continuous, restricting the decision breaks available to the forest. Additional methods that could be tested for improving this model are breaking continuous variables into buckets rather than binary variables, and further tuning the parameters for tree depth and number of observations required per leaf node.

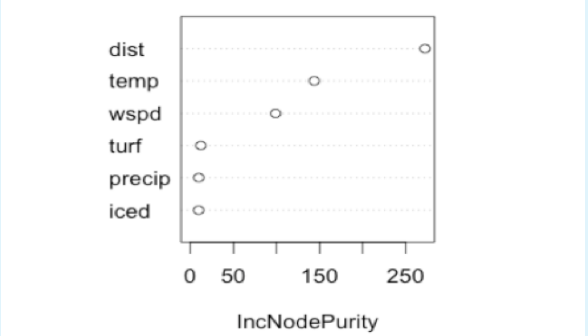

FIGURE 5: RANDOM FOREST FEATURE IMPORTANCES

Neural Networks

Neural Networks were the next type of model tested. This class of model loosely imitates the structure of neurons in the brain, beginning with one node per variable in the outer layer, a number of "hidden" nodes in the inner layers, and then an end node for the model's output. The model uses logistic activation functions to determine which nodes in the network will activate, and assigns weights between nodes that determine its output. The network uses a backpropogation algorithm to determine the optimal weights between its nodes, which allows it to improve the accuracy of its outputs. These models have gained a great deal of support in machine learning for their ability to gain deep insights on the patterns within data.

Although I tested deeper networks, this simple 7 x 3 x 1 network proved most effective:

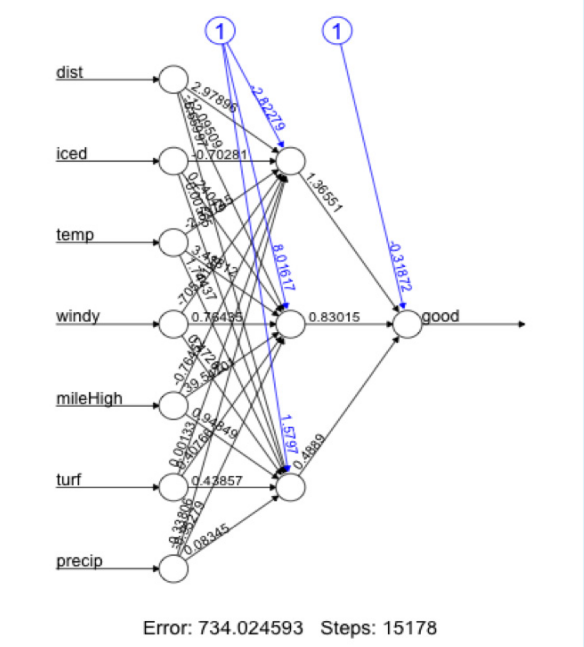

FIGURE 6: 7 X 3 X 1 NEURAL NETWORK ARCHITECTURE

While this was by far the most time consuming model to train and implement, it also yielded the best results of any individual model tested. This model's success was likely due to its allowance for flexible interactions between features. Whereas a linear model such as a logistic or linear

regression applies a single function to transform input features into outputs, the neural network has intermediary hidden layers that allow features to interact before being transformed into an output. The result is a model that learns its own features, as opposed to other models which are constrained to the features that a user feeds them. The slow learning of this model, however, poses a significant problem for testing and improvement, making parameter-tuning and model evaluation methods such as cross validation time consuming in the absence of improved computing resources.

Multiple Linear Regression

Finally, I tested a multiple linear regression model. Because of the logistic nature of the relationship between distance and field goal difficulty (the difference between the difficulty of a 55- and 65-yard field goal difficulty is much larger than that between a 25 and 35-yard attempt), simply throwing all the above variables into a standard MLR model will not fit the data well. Adding the log of the distance into a model with all of the previously used variables, however, yielded quality results, with an MSE measurement beating that of the random forest model and the original Clark logistic model. This MLR model allows for the non-distance variables to be linear intercept variables, and also makes for the highest ease of interpretation of all the models used, using the standard MLR model at the bottom of the page. Table 7 shows the regression output.

Different Models, Different Predictions

It should not come as a surprise that different models yield slightly different predictions. The regression outputs and error measurements mentioned up until this point in the paper, however, do not display model patterns well. Because distance is such a strong predictive factor in this particular set of models, simply showing probabilities graphed against distance seems to do the trick for illustrating probability patterns. The final models are displayed in Figures 7, 8, 9, and 10.

And the winner is...

While logistic regression is the standard for probabilistic measurements such as this, it was not the most precise of those tested. The best performing model in this situation is actually one that has not been mentioned yet. While the neural network performed best among all the models al-

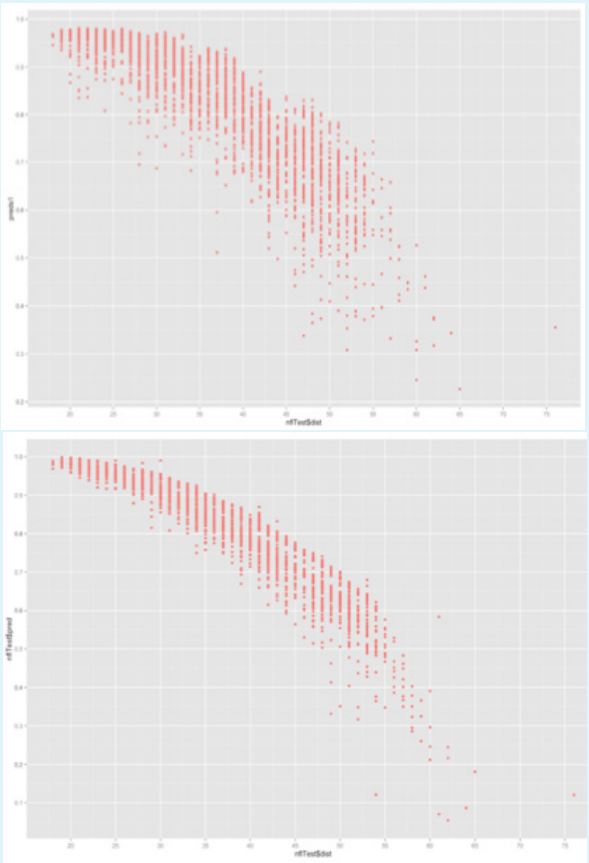

FIGURE 7 (TOP): RANDOM FOREST
FIGURE 8 (BOTTOM): NEURAL NETWORK

Variable	Coefficient	Std. Error	P-Value
Intercept	-0.3806	0.2146	0.0762
Distance (yards)	-0.0318	0.0024	0.0000
ln(distance)	0.6473	0.0850	0.0000
Temp (°F)	0.0008	0.0002	0.0001
Precipitation	-0.0398	0.0147	0.0069
Windy (≥ 10mph)	-0.0222	0.0076	0.0037
Turf	0.0417	0.0070	0.0000
Iced	-0.0272	0.0146	0.0629

TABLE 7: MLR MODEL OUTPUT

$$P(make) = \beta_0 + \beta_{dist}X_1 + \beta_{ln(dist)}X_2 + \beta_{cold}X_3 + \beta_{precip}X_4 + \beta_{windy}X_5 + \beta_{turf}X_6 + \beta_{alt}X_7 + \beta_{iced}X_8$$

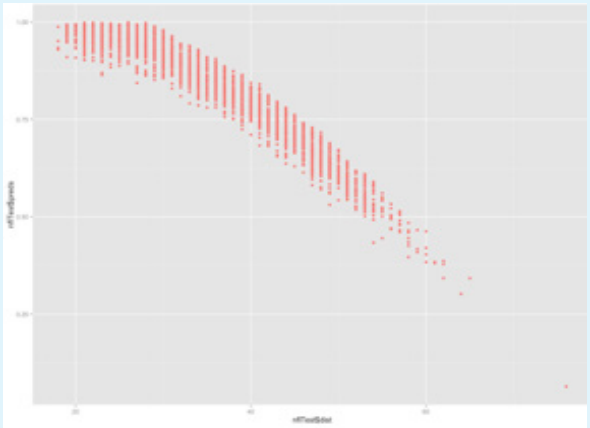

FIGURE 9: MULTIPLE LINEAR REGRESSION

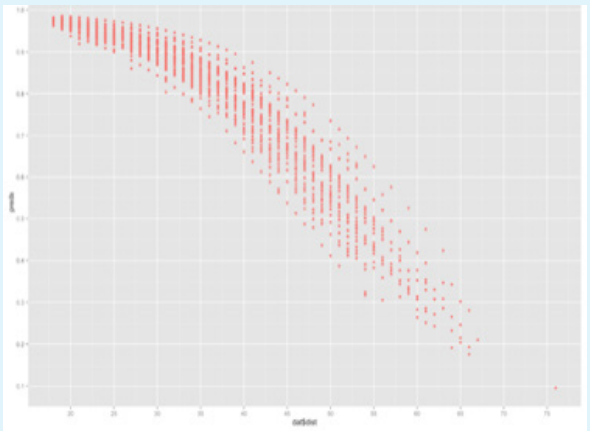

FIGURE 10: LOGISTIC REGRESSION

ready discussed, the best performer was one that combined the results of this and the second logistic regression model, taking the average of the predictions of the two. Table 8 shows how all the models compared in terms of MSEs (mean squared errors).

MSE is an error measurement taking the average squared distance between the predicted values (from the model) and their true values (1 if made, 0 if missed) across all predictions made. The smaller the MSE score, the more accurate the model. The error measurements in this case offer little descriptive significance as there is no true method for evaluating the accuracy of an observation's probability estimate. The scores serve merely to rank the models against one another.

While the neural network, MLR model, and logistic regression model with icing included were able to beat the logistic regression benchmark set by Clark, the best model of all was that which combined the predictions of the top two performers.

WHY DOES THIS MATTER?

I present two scenarios to illustrate how the analysis presented may be relevant to decision-making in NFL games.

The first took place on January 19, 2002, when the New England Patriots were playing the Oakland Raiders in the AFC divisional playoff game. The Patriots were down 13-10 when Adam Vinatieri lined up for a 45-yard attempt with time running low. In 19-degree weather and heavy snow, Vinatieri's kick went through the uprights with only 27 seconds left, and the Patriots went on to win the game in overtime on another Vinatieri field goal.

This snowy field goal has gone down in NFL history as one of the greatest clutch kicks of all time. It also happens to be an excellent example of one of the worst possible conditions to kick a field goal in, showing the need for models such as those discussed in this paper. The probabilities of Vinatieri's kick going in had the game been in Oakland vs. in New England are illustrated in Figure 11.

Had the game taken place in Oakland with its temperate climate, Vinatieri's 45-yard attempt would have had a very reasonable 71% likelihood of succeeding. In the New England blizzard, however, it had much riskier odds, at 54%. Though the kick was ultimately good, this model goes to show just how high the stakes were that Sunday afternoon, and the amount of pressure that Vinatieri, the coaching

Rank	Model	MSE
1	Neural Net and Logistic #2 Combined	0.1301
2	Neural Network	0.1302
3	Logistic #2 (icing included)	0.1303
4	Multiple Linear Regression	0.1303
5	Logistic #1 (Clark replication)	0.1317
6	Random Forest	0.1322

TABLE 8: MSE OF DIFFERNT MODELS

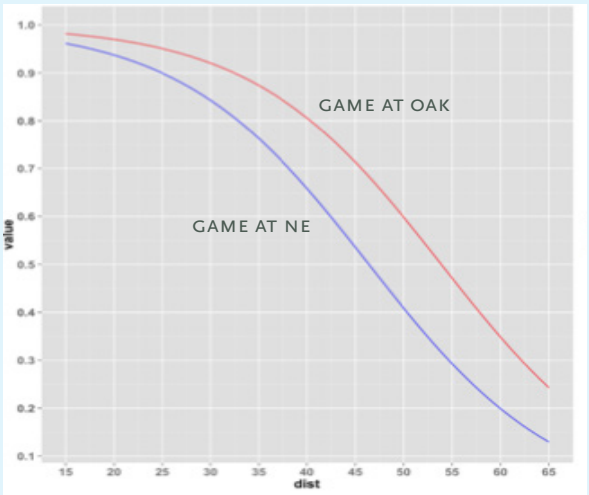

FIGURE 11: IMPACT OF WEATHER CONDITIONS ON THE SNOW BOWL GAME.

staff, and the New England fan base must have felt as they waited for the result.

The second situation is one that has not happened yet, but one that every NFL coach should be concerned about: a kicker's missed extra point costing his team the game. Moving the extra point back 13 yards may still seem like an easy conversion, being a 32-yard attempt, but as these models show, there is a noticeable decline in the probability of success in going from a 19 to a 32-yard kick. The result of this decreased likelihood was seen across the league this season, with extra point conversion percentages decreasing from 99.6 to roughly 95 percent with the implementation of the new rule. While nobody has lost a game due to this rule change just yet, this new area of risk is a perfect example of the importance of understanding field goal likelihoods for making effective coaching decisions. This increased risk will make two-point conversion attempts increasingly appropriate in the NFL.

As is seen in these two scenarios, understanding field goal conversion likelihoods is crucial to a coach's decision-making process. Understanding not only that distance, climate factors, and situational pressure have tangible impacts on field goal success likelihoods, but also to what extent they do, could make all the difference in weighing the risk and potential reward of a crucial coaching decision.

To conclude, this study has two main findings:

First, that field goals follow predictable patterns, allowing their probabilities of success to be modeled to a useful degree. This can and should be taken into consideration by coaching staffs during their evaluations of game time decisions.

Second, that good kickers really do matter. Finding a great one like Rob Bironas or Sebastian Janikowski can make all the difference in a 16-game season where playoff outcomes are not determined until the very last week, or in a playoff game where a single point-added above average replacement is all that stands between being a runner-up and a Super Bowl champion.

While baseball has been thoroughly transformed by sabermetrics and the Moneyball revolution, football still has a way to go in its acceptance of statistics as the governing dynamics of certain aspects of play, given its smaller sample sizes because of its 16-game seasons. Field goal kicking, however, being the isolated component of play that it is, has the potential to be one of the first great leaps in the quantification of the ways that coaches, scouts and fans view the game. Through the acceptance of methods and models such as those discussed in this paper, NFL organizations can make great strides in improving their abilities to make data-informed decisions and maximize the potential values of their rosters.

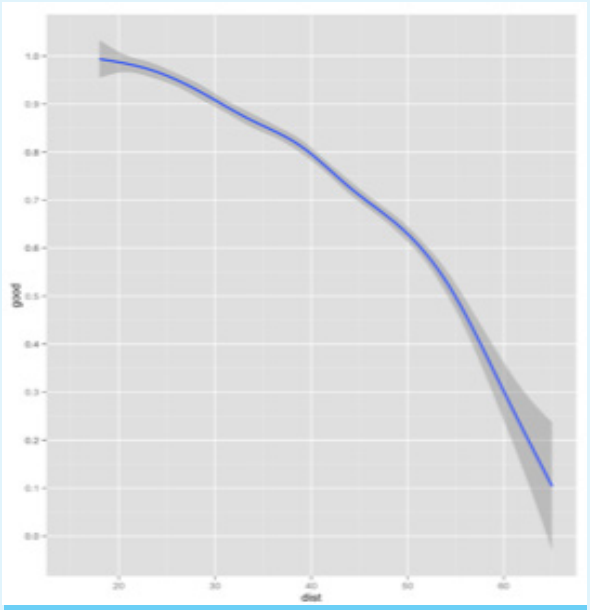

EXHIBIT A: LEAGUE-WIDE FG% BY DISTANCE, 2000-2013

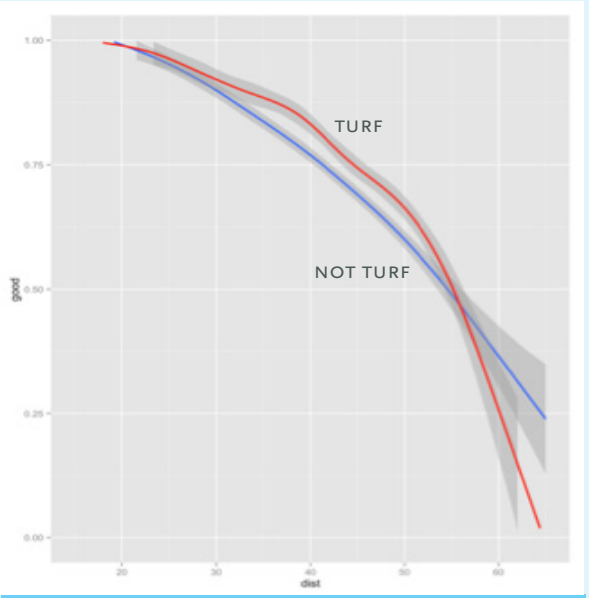

EXHIBIT B2: EFFECTS OF PARAMETER CHANGES — TURF EFFECT

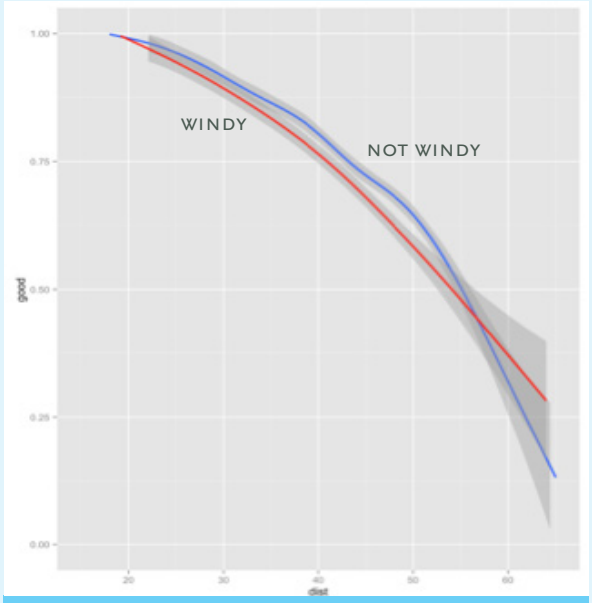

EXHIBIT B1: EFFECTS OF PARAMETER CHANGES — WIND SPEED EFFECT

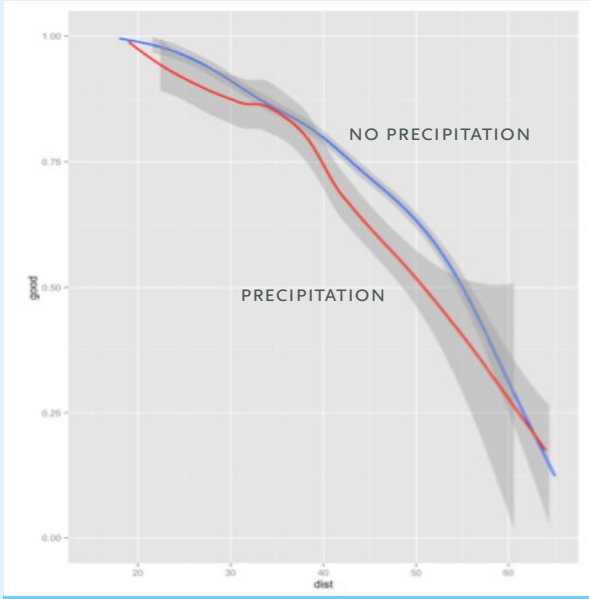

EXHIBIT B3: EFFECTS OF PARAMETER CHANGES — PRECIPITATION EFFECT

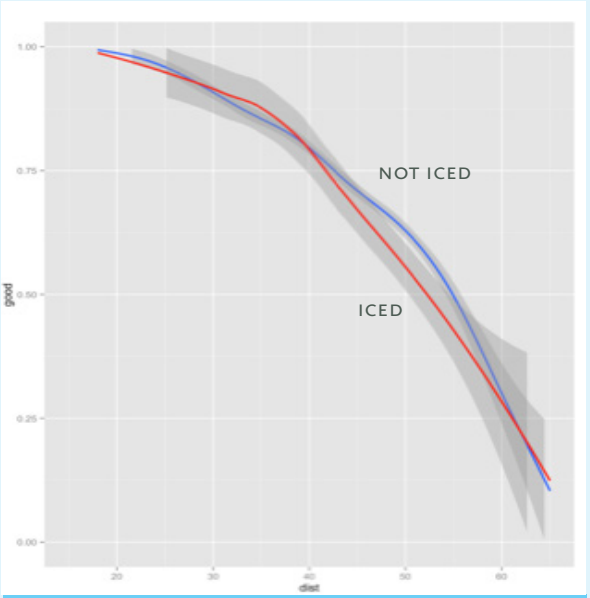

EXHIBIT B4: EFFECTS OF PARAMETER CHANGES — ICING THE KICKER

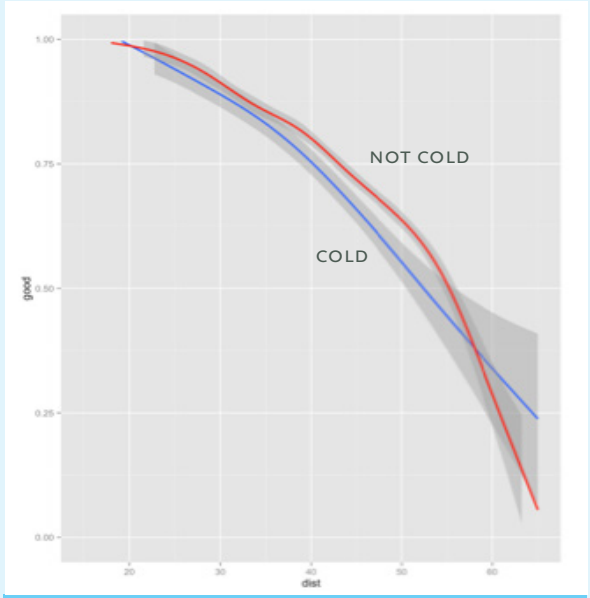

EXHIBIT B6: EFFECTS OF PARAMETER CHANGES — COLD WEATHER EFFECT

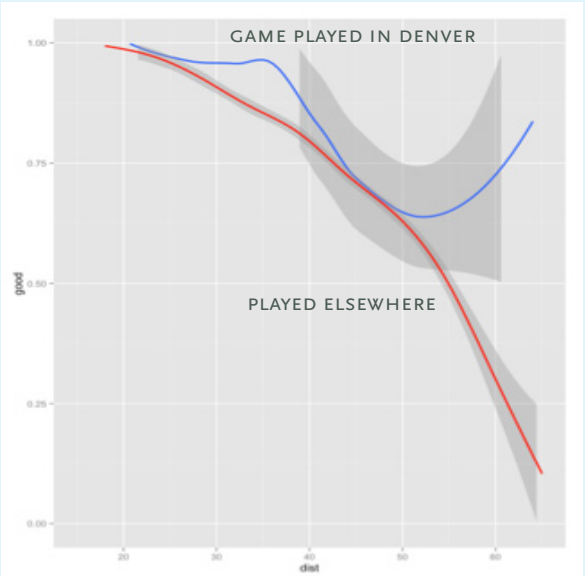

EXHIBIT B5: EFFECTS OF PARAMETER CHANGES — MILE-HIGH EFFECT

Variable	Coefficient	Std. Error	P-Value
Constant	5.953	0.220	<0.0005
Distance (yards)	-0.106	0.003	<0.0005
Cold (<50°F)	-0.341	0.061	<0.0005
Precipitation	-0.280	0.099	0.005
Windy (≥10 mph)	-0.140	0.055	0.011
Turf	0.299	0.053	<0.0005
Altitude (≥4000 ft.)	0.694	0.157	<0.0005

EXHIBIT C: ORIGINAL STUDY COEFFICIENTS.

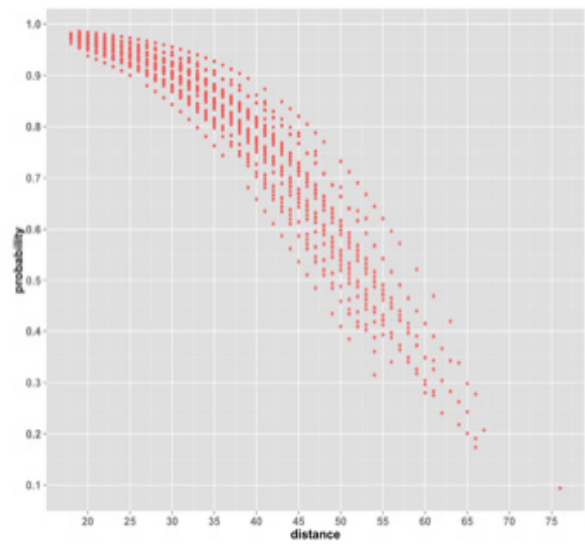

EXHIBIT D: PROBABILITY PREDICTIONS FROM LOGISTIC MODEL 2

ENDNOTES

1. Torin, Johnson, and Stimpson, "Going for Three: Predicting the Likelihood of Field Goal Success with Logistic Regression."
2. *Affordable NFL Data Plans*, Armchair Analysis.
3. Brian, "Altitude and Field Goals."
4. Torin, Johnson, and Stimpson, "Going for Three: Predicting the Likelihood of Field Goal Success with Logistic Regression." *NFL.com.*,"Stephen Gostkowski."
5. Matt, "A Step by Step Backpropagation Example," March 17, 2015.

REFERENCES

Clark, Torin, Aaron Johnson, and Alexander Stimpson. "Going for Three: Predicting the Likelihood of Field Goal Success with Logistic Regression." MIT Sloan Sports Analytics Conference (2013). Accessed January 11, 2016.

Affordable NFL Data Plans. Accessed 11 January 2016. Armchair Analysis.

Burke, Brian. "Altitude and Field Goals." Advanced Football Analytics. Accessed January 11, 2016.

"Stephen Gostkowski." NFL.com. Accessed January 11, 2016.

Mazur, Matt. "A Step by Step Backpropagation Example." March 17, 2015. Accessed January 11, 2016.