

Analyzing Factors That Affect NFL Kicking Performance

Cornell Sports Analytics Club

By Duncan Park, Sameer Nathawat, and
Antonio Ramirez

Introduction:

NFL kickers are often placed in high-pressure situations where small factors can dramatically affect success. Our project investigates whether situational variables impact field-goal accuracy.

Our goal is to understand which factors truly matter and how teams can use this information to make more strategic in-game decisions.



Variables Analyzed:



Home vs. Away



Distance



Score Differential



Icing the Kicker



Enclosed vs. Open



Wind

Methodology:

We used **NFLFastR**, an R package that provides play-by-play data including factors like kick distance, home or away, wind speed, score differential, roof or no roof, and icing the kicker (if an opponent called a timeout before the kick).

We filtered the data to include **field-goal attempts only** (extra points excluded) for the **past three NFL seasons (2022-2024)**.

Using R, we created a dataset based on the variables we were analyzing. We set three outcomes for the kick: **made, missed, or blocked**.

After cleaning the dataset, we analyzed correlations between field-goal outcomes and each variable. Finally, we weighted the variables and built models to estimate the effects of these variables on whether a field goal is made, missed, or blocked based on historical sets of NFL data.

Random Forest Classifier:

A **random forest classifier** is a type of machine learning model that uses many **decision trees** to make predictions.

We created two types of random forest classifiers:

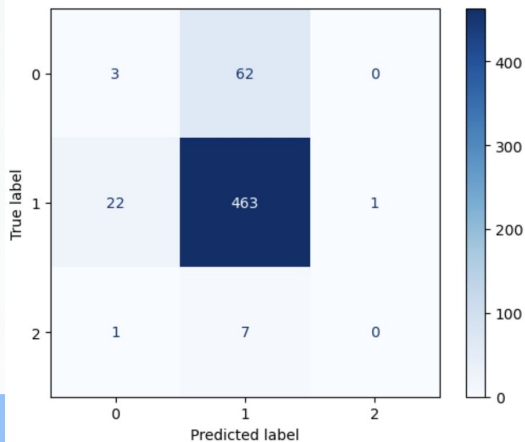
Baseline Model: Treated all kick outcomes equally when training the model.

Weighted Model: Due to misses and blocks being a small share of all kick outcomes, we oversampled these minority classes so the model could better learn their patterns. We used the RandomOverSampler function from the sklearn library to perform this.

Random Forest Classifier:

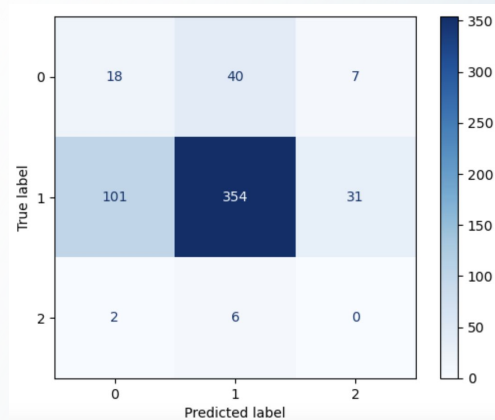
Baseline Model:

	precision	recall	f1-score	support
0	0.12	0.05	0.07	65
1	0.87	0.95	0.91	486
2	0.00	0.00	0.00	8
accuracy			0.83	559
macro avg	0.33	0.33	0.33	559
weighted avg	0.77	0.83	0.80	559



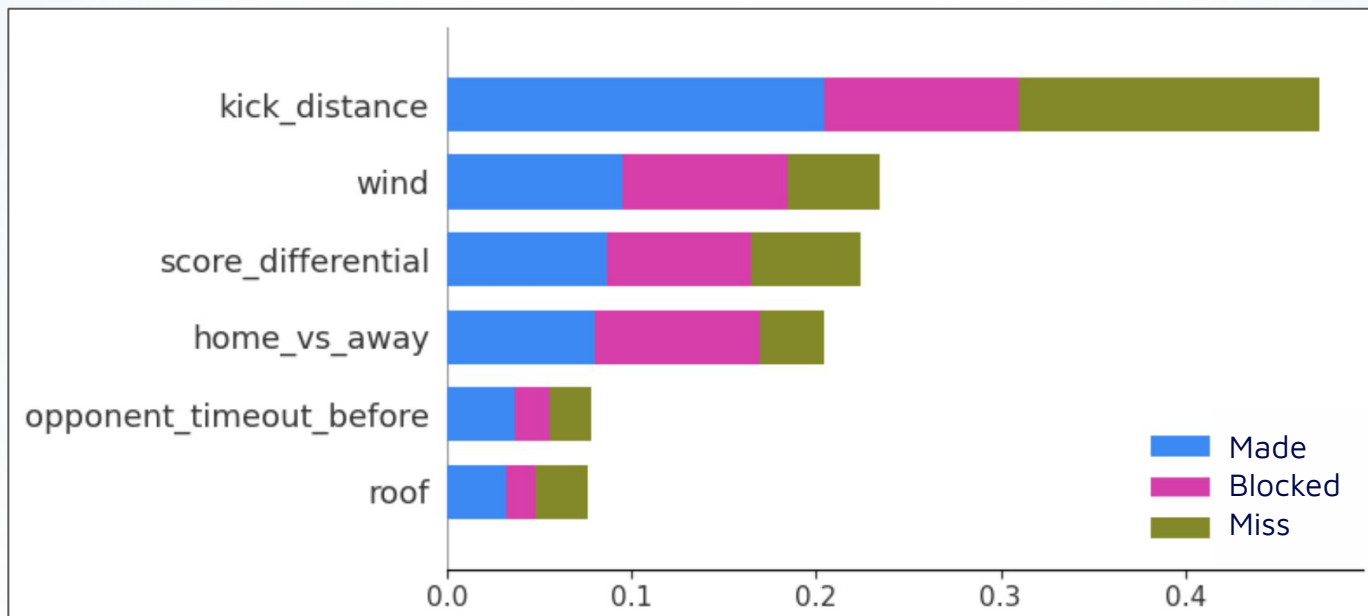
Weighted Model:

	precision	recall	f1-score	support
0	0.16	0.65	0.26	65
1	0.95	0.38	0.55	486
2	0.00	0.00	0.00	8
accuracy			0.41	559
macro avg	0.37	0.34	0.27	559
weighted avg	0.84	0.41	0.51	559



Random Forest Classifier:

Feature Importance:



Multinomial Logistic Regression Model and Confusion Matrix

Our **multinomial logistic regression model** allows us to model all three outcomes of a field goal (made, missed, blocked) simultaneously. We set the **baseline as “missed,”** which allowed for our model to show how different variables change the odds of a kick being made rather than missed.

Each coefficient of the variables we analyzed reflects the variables importance in the likelihood of kick success (assuming all other factors constant).

We had to **upsample the data** in order to ensure the size of the missed and blocked kicks data wasn't too small and prevent our model from picking “made” every time.

Multinomial Logistic Regression Model

Class	Predictor	Coefficient	StdError	z_score	p_value	Significant
blocked	(Intercept)	0.222864676	0.181541056	1.2276269	2.195871e-01	No
blocked	kick_distance	-0.026638464	0.003494505	-7.6229569	2.486900e-14	Yes
blocked	wind	-0.026486754	0.008466923	-3.1282622	1.758432e-03	Yes
blocked	roofoutdoors	1.021679370	0.097211271	10.5098859	0.000000e+00	Yes
blocked	home_vs_awayhome	0.812004411	0.062470347	12.9982376	0.000000e+00	Yes
blocked	score_differential	-0.016760019	0.003570480	-4.6940516	2.678465e-06	Yes
blocked	opponent_timeout_before1	-0.101225576	0.078429299	-1.2906602	1.968215e-01	No
made	(Intercept)	4.952375354	0.174064120	28.4514429	0.000000e+00	Yes
made	kick_distance	-0.111218665	0.003618765	-30.7338720	0.000000e+00	Yes
made	wind	-0.041277982	0.009426408	-4.3789726	1.192401e-05	Yes
made	roofoutdoors	-0.076233014	0.102559855	-0.7433027	4.572985e-01	No
made	home_vs_awayhome	0.328273172	0.065589137	5.0049930	5.586405e-07	Yes
made	score_differential	0.004739329	0.003792081	1.2497963	2.113740e-01	No
made	opponent_timeout_before1	-0.403110484	0.087518915	-4.6059813	4.105253e-06	Yes

Note: Variables are meaningful if:

$|Z\text{-score}| = | \text{coef} / \text{standard error} | > 2$

Confusion Matrix

Confusion Matrix and Statistics

	Reference		
Prediction	missed	blocked	made
missed	1188	780	557
blocked	734	1019	405
made	453	576	1413

Accuracy : 0.5081

	Class: missed	Class: blocked	Class: made
Sensitivity	0.5002	0.4291	0.5949
Specificity	0.7185	0.7602	0.7834

Interpreting the Matrix:

- Out of the missed kicks, **1188 were correctly predicted to be missed**, 734 blocked, and 453 made
- Out of the blocked kicks, **1019 were correctly predicted to be blocked**, 780 missed, and 576 made
- Out of the made kicks, **1413 were correctly predicted to be made**, 557 missed, and 405 blocked

Statistical Analysis:

- Model is accurate ~**50% of the time**
 - **Better than guessing**, which would be 33.3%
- Model is **best at predicting when a field goal would be made** (~60% accurate), second best at missed, and worst at blocked
 - Makes sense logically when thinking about blocked field goals

Conclusion:

Ultimately, we found that kicking distance was the biggest predictor of the success of a field goal.

While other variables such as score differential, home versus away, enclosed versus open stadiums, and wind play a factor in kicking, their effects are smaller.

To continue our work, we could expand the number of years analyzed or more deeply investigate other factors such as humidity, field conditions, and kicker biomechanics.

We also could discard “blocked” kicks as they are the hardest to predict and are more random, and instead turn our project solely to made vs. missed kicks.



Thank you! Questions?