

# Field Goal Difficulty Model: Estimating Success Probability

## Purpose

This model is built to estimate how difficult a field goal attempt is based purely on physical conditions. It is not about predicting whether a specific kicker will make the kick, but about understanding how tough the attempt is in general. By using only objective inputs like distance, wind, temperature, and surface type, the model stays clear of bias from things like game context or a kicker’s track record. The goal is to give decision-makers a clean, reliable baseline to support scouting reports, pregame planning, and quick decisions on the sideline.

## Data & Feature Engineering

I used play-by-play data from nflfastR and filtered for valid field goal attempts. Weather conditions were cleaned up and binned into simple categories based on accuracy trends and football logic. For example, temperatures below 50°F were labeled “Cold” because make rates noticeably dropped in that range. I excluded game context variables like quarter, score, or playoff status to keep the focus strictly on kick difficulty. I also added interaction terms like wind speed times distance and cold times wind to reflect how certain factors matter more under specific conditions, such as wind having a bigger impact on longer kicks.

## Modeling Approach

- **Model:** Logistic Regression with SMOTE and interaction terms
- **Target:** Binary outcome (Make/Miss), base rate ~84%
- **Key Inputs:** Kick distance, weather (binned temperature and wind), surface type, altitude, precipitation, humidity, and interaction terms like wind speed × kick distance and cold × wind
- **Calibration:** Applied Platt scaling to improve probability accuracy
- **Threshold:** 0.50 selected for balance between sensitivity and specificity

I tested XGBoost and a few other models, but logistic regression was the better fit and the most interpretable. It provided clearer insight into which factors mattered and did a better job balancing made and missed kicks during training. I used SMOTE to address the imbalance by generating more missed kick examples. After training, I applied Platt scaling to calibrate the predicted probabilities, making sure the model's confidence matched real-world outcomes. XGBoost was well-calibrated overall, but it leaned too heavily toward predicting every kick as a make, which made it less useful for spotting the tougher attempts.

## Model Takeaways

Model	AUC	Brier	LogLoss	Sens	Spec	Balanced Acc
XGBoost	0.765	0.135	0.481	0.989	0.057	0.523
Logistic (SMOTE)	0.771	0.202	0.583	0.666	0.736	0.701

The model achieved an AUC of 0.77 (a measure of how well the model separates made and missed kicks), showing strong ability to rank kicks from easier to harder. Brier Score dropped from 0.2041 to 0.1148 after calibration (lower is better), meaning the predicted probabilities better matched actual outcomes. Logistic regression outperformed XGBoost in balanced accuracy, handling both makes and misses more effectively and avoiding the tendency to overpredict makes. The model is well-calibrated, easy to interpret, and consistent with football logic. Kick distance, wind, and precipitation reduced success probability. Interaction terms like wind times distance and cold times wind also had strong negative effects. Cold and night games showed small positive effects, which may reflect better preparation or selective usage by teams.

## Model Scope & Use

This model is designed to serve as a baseline tool for evaluating field goal difficulty, whether it is for scouting reports, pregame planning, or quick decisions on the sideline. It intentionally leaves out kicker-specific traits, game state, and execution issues like snap or hold quality to keep the model clean, flexible, and focused on difficulty alone. These factors can always be layered in later if the goal shifts toward explanation, such as understanding why a kick missed rather than how difficult the attempt was. At its core, the model is built to answer one question clearly: How hard is this kick? It can also help inform 4th-down decisions and support kicker evaluation by providing an objective view of kick difficulty across situations.