SuccessVision: Modeling Offensive Play Success from Pre-Snap Information

3rd and 4th Down Pre-Snap Prediction in the NFL

Team 8 • EAS 508 • Colin Seiler • Aidan Collins • Nayan Paliwal • Abdul Salam

Introduction

This study examines whether offensive play success in the NFL can be predicted using only pre-snap information. Using play-by-play data from the 2016–2023 seasons, we frame success prediction as a binary classification problem and evaluate multiple machine learning models using features derived from game context, formations, and personnel. While predictive strength varies by down, performance improves substantially on third and fourth downs, with yards-to-go emerging as the dominant predictor of success.

Data Sets & Features

We construct a pre-snap feature set capturing game context, field position, formations, and personnel, including down, time remaining, yards-to-go, score differential, formation markers, and offensive and defensive position counts. To incorporate team and player strength, we include Madden player ratings for offensive skill positions as a consistent proxy for player ability, assigning default values when ratings are missing. Additional team-level metrics from the previous season are also included to expand contextual information while maintaining a strictly pre-snap framework.

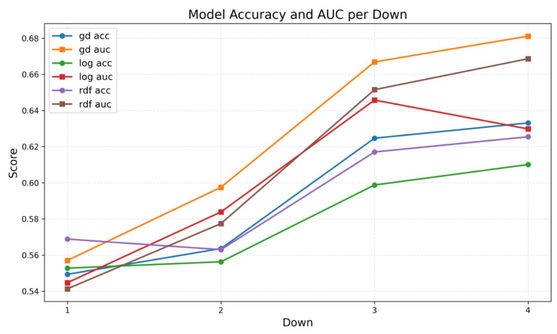
Methods

We train Logistic Regression, Random Forest, and Gradient Boosting models using scikit-learn. Logistic Regression provides a baseline; Random Forest uses bagged trees; Gradient Boosting builds sequential trees to reduce loss. Because success is imbalanced, we evaluate using Precision, Recall, F Score, and especially AUC. Accuracy alone is insufficient, so we emphasize metrics that reflect classification quality.

Results (All Downs)

When evaluated across all downs, model performance does not exceed the baseline accuracy of 0.60, though all approaches achieve AUC values slightly above 0.5, indicating some non-random structure in the predictions. These results suggest that evaluating plays collectively across all downs obscures situational differences in play behavior, motivating a focused analysis of more constrained scenarios on third and fourth downs.

Model	Accuracy	AUC
Baseline	0.600	0.500
Logistic	0.560	0.586
RandForest	0.580	0.585
XGBoost	0.570	0.604



Final Down Results

Restricting the analysis to third and fourth downs leads to substantially improved model performance compared to all-down results. All models outperform the baseline, with the ensemble achieving the best overall performance (Precision = 0.61, Accuracy = 0.64, AUC = 0.67). Yards-to-go consistently emerges as the dominant predictor, indicating that success on final downs is largely driven by distance-based constraints.

Evaluation Metrics from Final Downs

XGBoost

	Precision	Recall	F1-score	Support
0.0	0.69	0.64	0.66	4725
1.0	0.56	0.61	0.58	3534
Acc.			0.63	8259
Macro Avg	0.62	0.63	0.62	8259
Weighted Avg	0.63	0.63	0.63	8259

Logistic Regression

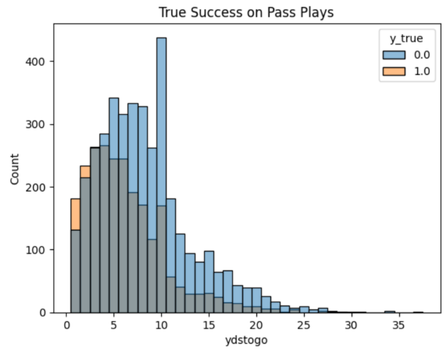
	Precision	Recall	F1-score	Support
0.0	0.69	0.62	0.65	4725
1.0	0.55	0.62	0.58	3534
Acc.			0.62	8259
Macro Avg	0.62	0.62	0.62	8259
Weighted Avg	0.63	0.62	0.62	8259

Logistic Regression + Random Forest

	Precision	Recall	F1-score	Support
0.0	0.68	0.69	0.68	4725
1.0	0.57	0.56	0.57	3534
Acc.			0.63	8259
Macro Avg	0.62	0.62	0.62	8259
Weighted Avg	0.63	0.63	0.63	8259

Discussion

The stronger performance on final downs reflects the more deterministic nature of third- and fourth-down situations compared to earlier downs. The dominance of yards-to-go suggests that while the models capture core structural relationships, additional features are needed to better explain more complex outcomes, particularly for passing plays.



Conclusion

Pre-snap machine learning models do not reliably predict offensive play success across all downs but provide meaningful predictive value on third and fourth downs, where play objectives are more constrained. Differences between rushing and passing plays, along with the dominance of yards-to-go, highlight the limits of distance-based features. These results show that pre-snap information is most actionable in high-leverage situations.

University at Buffalo The State University of New York

Engineering Science (Data Science) MS
School of Engineering and Applied Sciences
buffalo.edu