



Predicting NFL Run vs Pass Plays Using Machine Learning

Michael Gomes – UF DCP4300 – Fall 2025

Problem Statement

NFL offenses show tendencies based on game situation. This project predicts whether a play will be a **RUN** or a **PASS** before the snap using structured game data. Understanding these tendencies helps with coaching strategy, defensive planning, and football analytics.

Goal: Build a model that predicts RUN vs PASS using pre-snap features.

Why this matters:

- Play calling heavily affects game outcomes
- Defensive coordinators try to read tendencies
- Predictability changes based on down, distance, and score
- Machine learning can reveal hidden patterns

Dataset Overview

Source: NFL Play-by-Play 2009–2018 (Kaggle)

Filtered (RUN/PASS only): 49,267 plays

Features Used:

- Down
- Distance (ydstogo)
- Field Position (yardline_100)
- Quarter
- Score Differential
- Offense Team (posteam)
- Defense Team (defteam)
- Home/Away (posteam_type)
- Game Year

Input

down	ydstogo	yardline_100	qtr	score_differential	posteam	defteam	posteam_type	game_date
1.0	10	56.0	1	0.0	PIY	TEN	home	2009-09-10
2.0	5	53.0	1	0.0	PIY	TEN	home	2009-09-10
3.0	8	56.0	1	0.0	PIY	TEN	home	2009-09-10
1.0	10	56.0	1	0.0	TEN	PIY	away	2009-09-10
2.0	10	56.0	1	0.0	TEN	PIY	away	2009-09-10

Output

play_type
pass
run
pass
run
pass

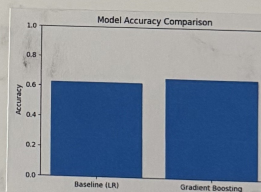
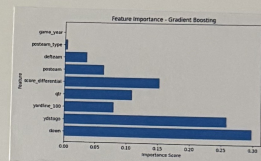
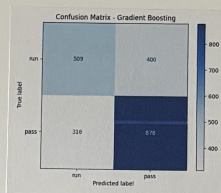
Results

Model Performance

- **Logistic Regression:** ~64% accuracy
- **Gradient Boosting:** ~68% accuracy
- Gradient Boosting outperforms baseline by ~4%

Key Insights

- PASS plays are slightly easier to predict than RUN
- Most important features:
 1. Down
 2. Yards to go
 3. Yardline_100
 4. Score differential
- Some play types overlap → play calling is not completely predictable



Methods

Two models were tested:

Baseline Model

Logistic Regression

- Simple, interpretable
- Establishes benchmark accuracy

Main Model

Gradient Boosting Classifier

- Handles non-linear patterns
- Better performance

Preprocessing

- Label encoding (RUN = 0, PASS = 1)
- Categorical encoding (teams, home/away)
- Train/test split (80/20)

Conclusion

- ML models can predict NFL play type with solid accuracy (~66–68%)
- Game situation variables have strong impact on play calling
- Gradient Boosting improves accuracy over a simple baseline

Future Work

- Add personnel groupings and formations
- Use team-specific models (each team behaves differently)
- Test deep learning or sequential models (RNN/LSTM)
- Include weather, field conditions, and coaching tendencies

GitHub QR Code



<https://github.com/michaelg133/NFL-Run-vs-Pass-Prediction>