Predicting pass or run plays in the NFL

By TJ and James

Business Question

Can we predict whether a team is going to run or pass the based in any given moment of a football game?

About the dataset

- The dataset contains details from 2,526 games across 10 seasons.
- There were 449,371* plays in those games.
- The objective of this project was to use information included in the dataset to predict if a team would call a run play (132,692 plays) or a pass play (186,677 plays).

*special teams plays were filtered out from the dataset (so only run/pass plays were used in the analysis)

Model Selection

Model	Accuracy Score
Random Forest	66.8
Decision Tree	69.0
XGBoost	70.2

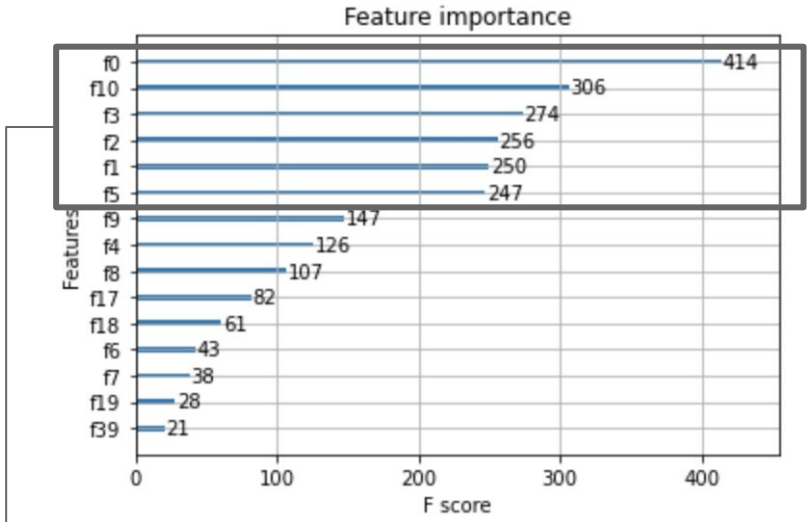
We performed 3 Models - Random Forest, Decision Tree, XGBoost to predict the playcall.

The model accuracy scores shown a slight increase with the most accurate model being the XGBoost.

Insights from analysis

Many attributes were considered in the predictive model, but the most important when it came to predicting run v pass were:

1. Yard line
2. Score differential
3. Time remaining
4. Yards remaining for first down



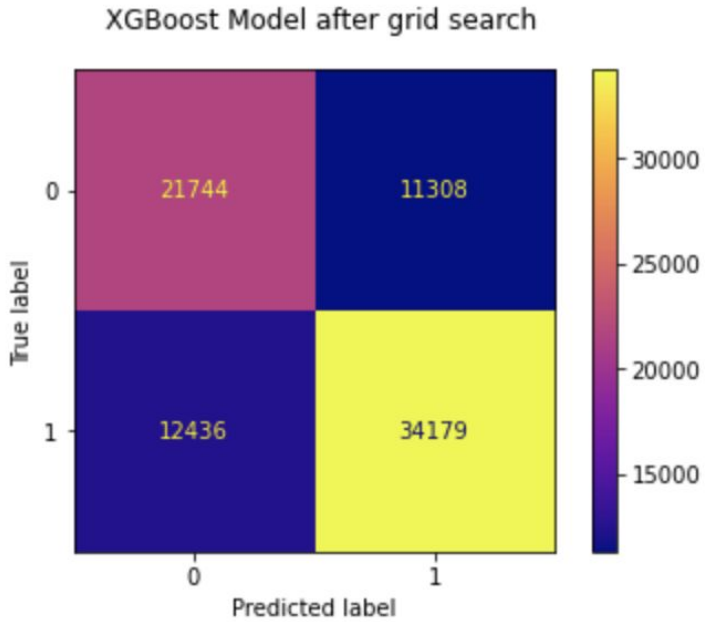
f0 = distance from endzone (yards)
f10 = offense score - defense score
f3 = time remaining in game (seconds)
f2 = time remaining in half (seconds)
f1 = time remaining in quarter (seconds)
f5 = distance from first down (yards)

Overview of the final model

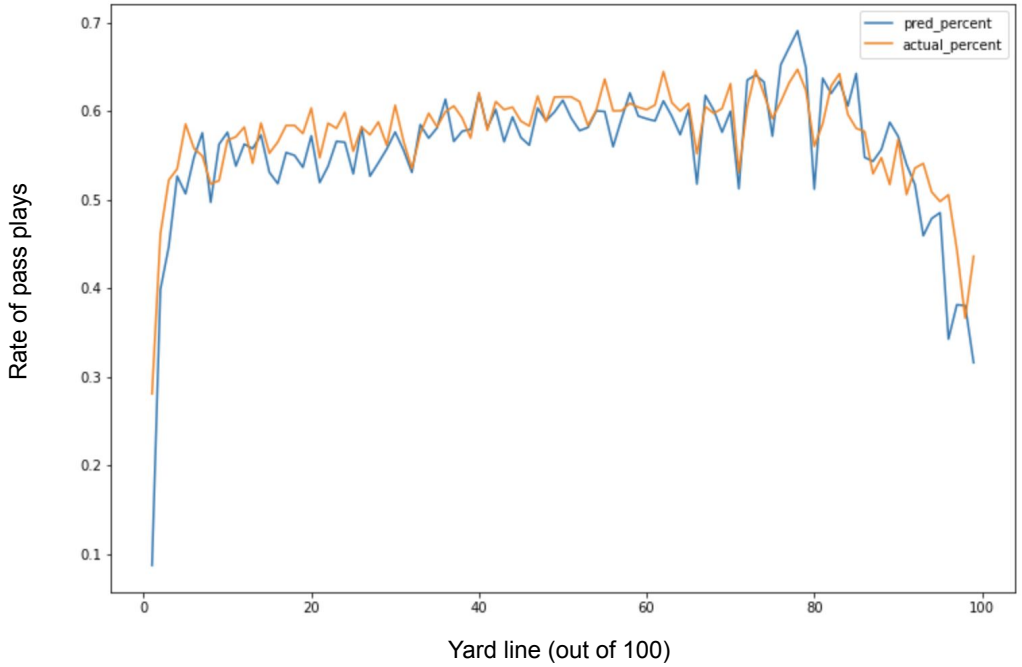
The model with the best performance was **XGBoost** and the optimal hyperparameters were determined using **grid search**

Performance on test data compared to dummy model (predicting pass on every play):

- Dummy accuracy - 58.5%
- XGBoost accuracy - 70.2%

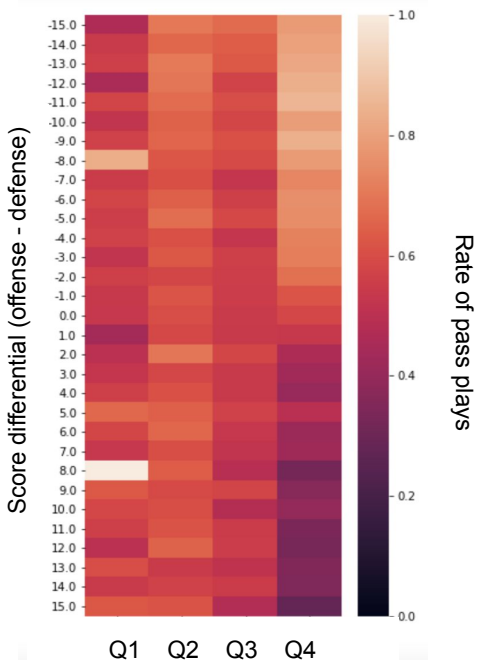


Pass percentage by field position



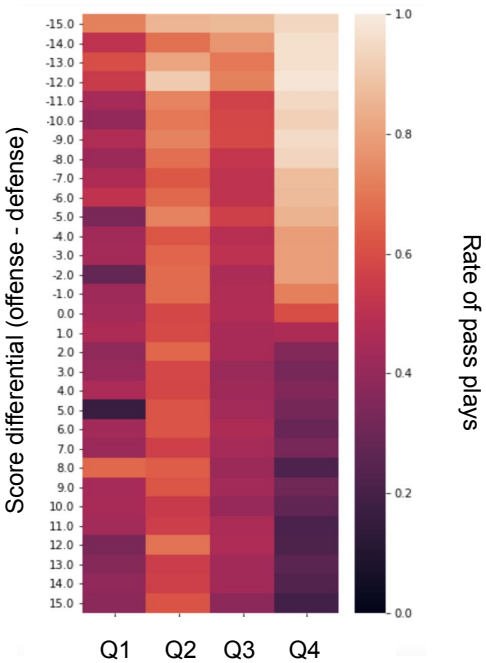
Predictions on the test data demonstrated similar trends to the actual percentage of pass plays called.

Pass % by quarter and score differential



Actual vs Predicted

Visual depiction of passing percentage by quarter and score demonstrate clear trends (e.g. teams are more likely to pass in Q2 than in Q1 and Q3; teams that are losing in Q4 are much more likely to pass than those that are winning).



Note: Score differentials of >14 were rounded to 15 to limit size of visual.

Next steps

What we can do to further improve our predictive model:

- Include more detailed features regarding players (including injuries, team changes, etc.)

- Create a play call tendency feature based on what plays a particular team has called earlier in the game