Predicting NFL Offensive Plays

Justin Knight

Problem

NFL coaches and general managers are woefully behind many other leagues in their utilization of advanced analytics to enhance in-game decisions

Here, I sought to give teams a leg up on the competition by building a model to predict the opponent's next offensive play

Binary classification - Play Type (pass vs rush)

Multiclass classification - Play Direction (left vs middle vs right)

Dataset

Postgresql relational database containing play-by-play, game scores, player stats, schedules, and much more from the 2009-2016 seasons and playoffs.

8 tables in total in the database

Clean Data

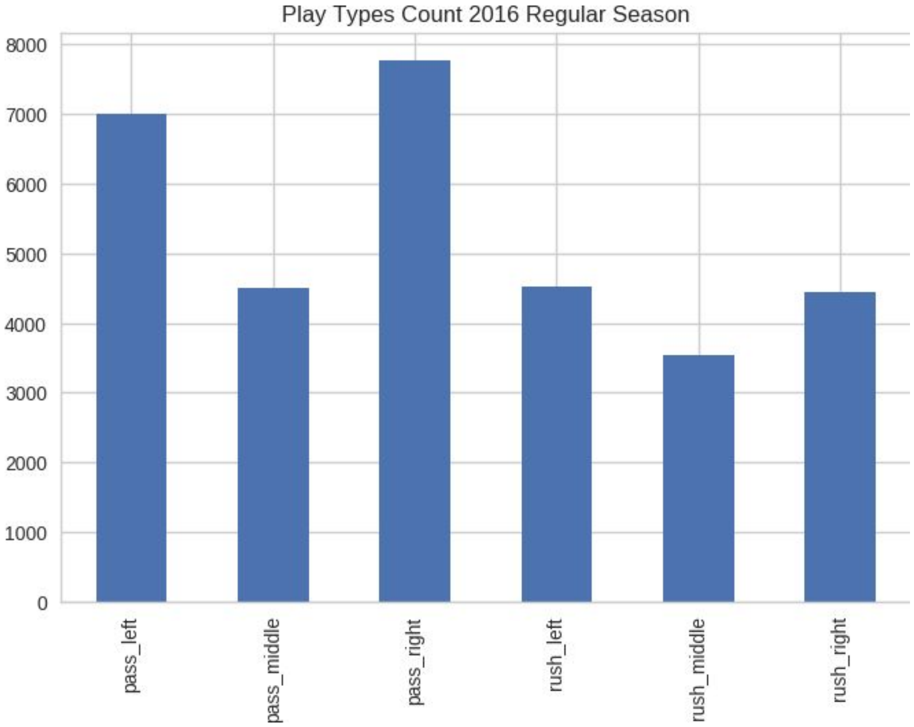
Wrangled and cleaned data

Included: integers, floats, and dummy encoded categorical variables

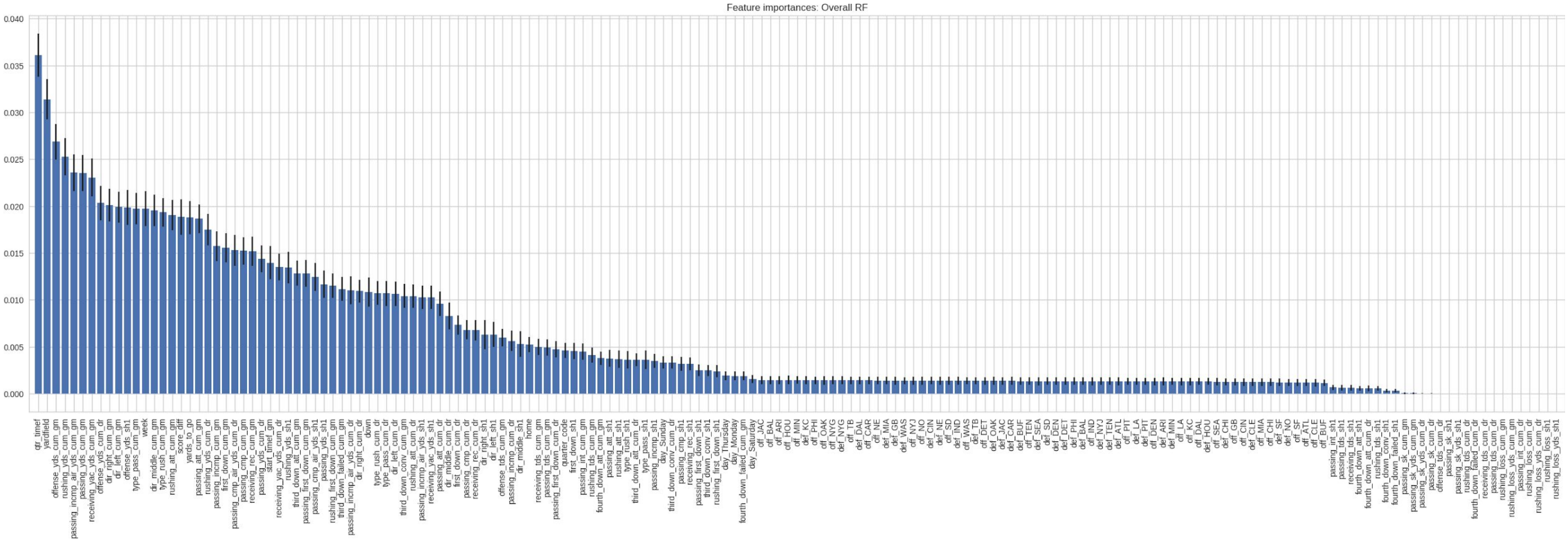
~32,000 plays

~170 features

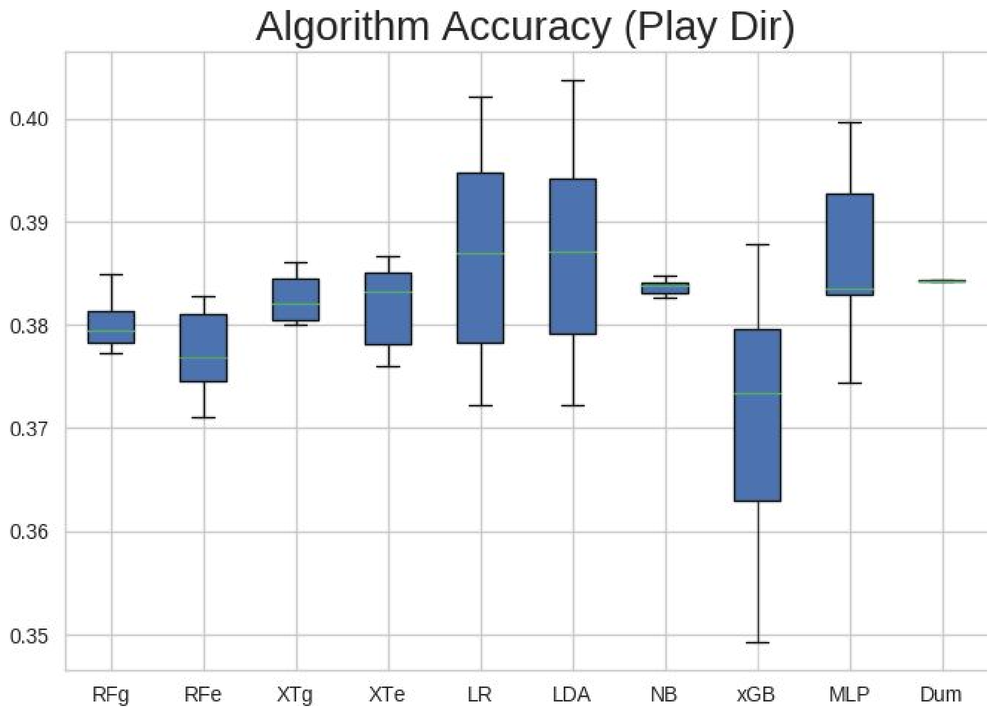
2016 Play Counts



Feature Importances: Play Direction

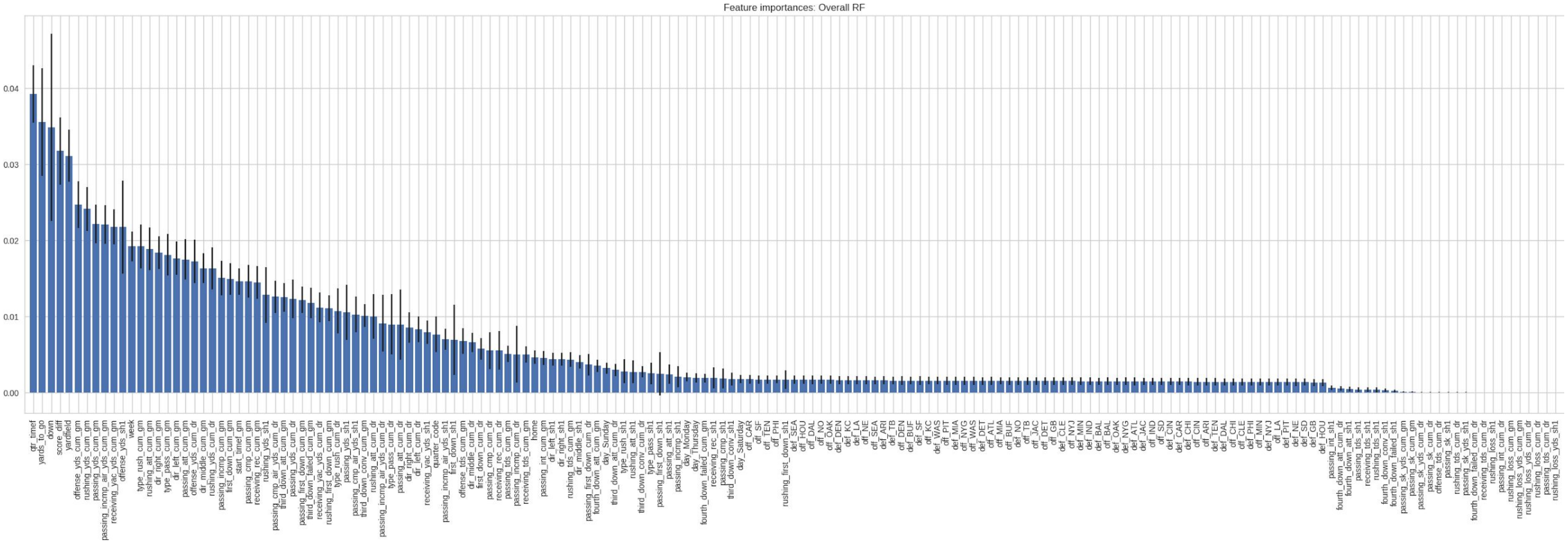


Machine Learning Algorithm Accuracy (Play Direction)

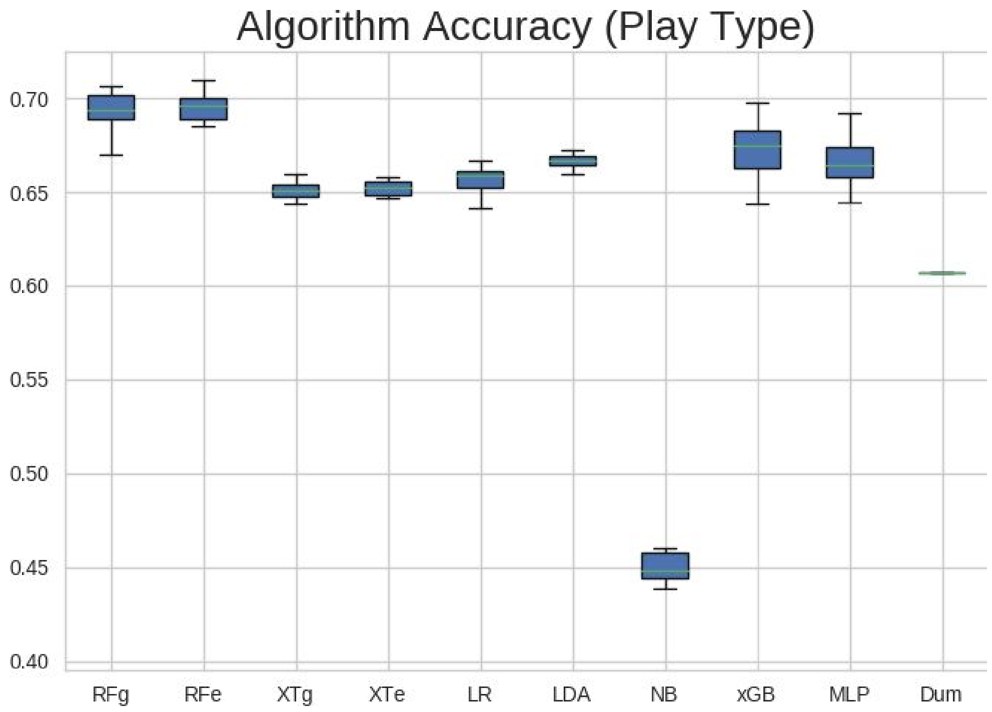


No model predictions for play direction performs significantly above labeling each example with the most frequent class (Dum), so the remainder of analyses will focus on play type

Feature Importances: Play Type



Machine Learning Algorithm Accuracy (Play Type)



Algorithm Ensemble

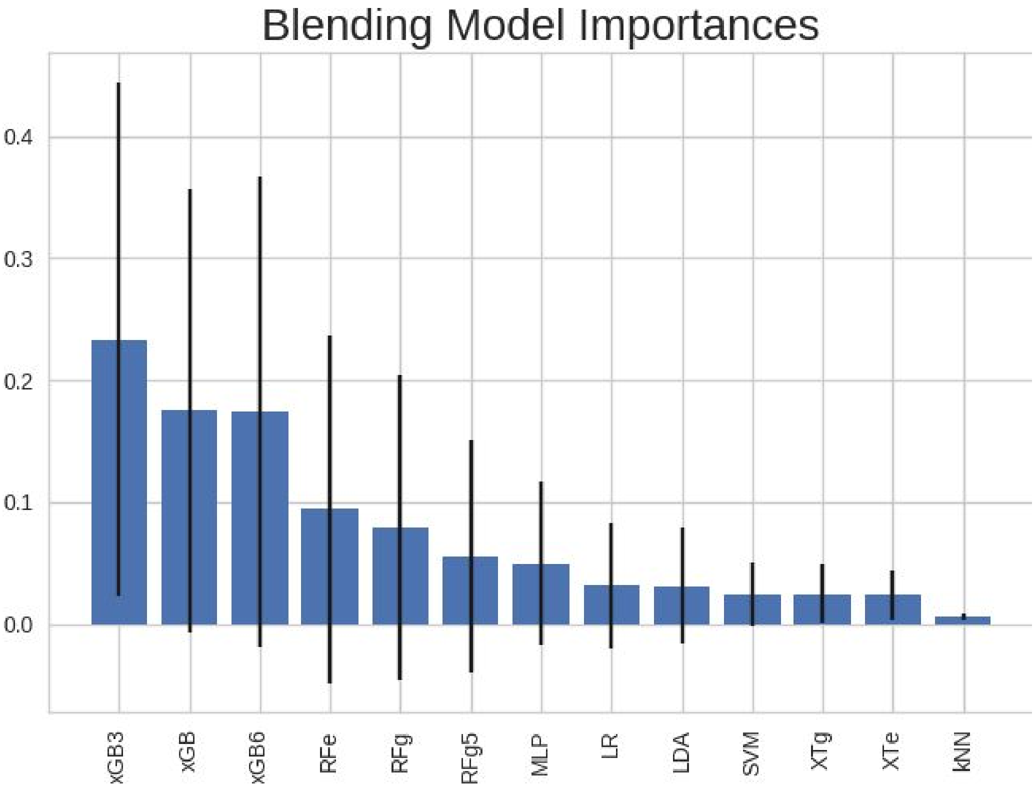
Ensembled the individual models together to enhance predictive power by capitalizing on the diverse errors (residuals) across the models

Predictive probabilities from the base models are then input as the training features for the ensemble blender model which was a Random Forest

Produces more informed and accurate predictions

73% accuracy

Blender Ensemble Contributions



Results and Recommendations

The best base models, Random Forest and Gradient Boosting, cannot compete with the final blending ensemble model

First, scheme your passing defense to be strongest outside, as plays over the middle are less frequent for both passes and rushes. For example, spread out the linebackers slightly more to give them better outside coverage.

Second, do not let the previous play overly dictate your defensive calls because overall game patterns are much more predictive of the next play, according to the random forest feature importances.

Third, on 2nd and 3rd use pass defenses far more frequently, as offenses are significantly more likely to call pass plays on later downs.

Future Directions

Train and test a recurrent neural network (RNN) with long short term memory (LSTM) and an attentional component to not only capture the most recent plays but also plays from previous drives and games in similar situations.

Collect additional data that includes features which are related to play direction.

Some examples include gameday weather, stadium cardinal direction, field surface, offensive line players and rankings, defensive and offensive team rankings, player rankings, and coaches tendencies.