

Predicting NFL play calls



Determining which teams are the most predictable and what determines their predictability



Mandatory Credit: Jayne Kamin-Oncea-USA TODAY Sports

Background

- NFL Play callers have a massive role in determining the outcome of individual plays
- Teams choose plays based on a number of factors
 - Score, personnel, strength of the opponent, formations
- If one team knows what play the other team is going to run, that team can put the necessary people in place to thwart the intended outcome for the other team
- As a result, the best offensive coaches are believed to have 'unpredictable offenses' that never let the defense know what's coming
- The sports analytics world is constantly improving in its scope and teams are always looking to get any edge they can

Problem Statement

- Is it true that the better NFL offensive coaches are less predictable than other coaches?
 - Does it matter if you're predictable if you're simply playing to your strengths, advantages?
- Which types of teams are more predictable: pass heavy teams or run heavy teams?
- For each team, what are the deciding factors in determining the play call?
- Across the league, which teams are affected by certain factors and not others?

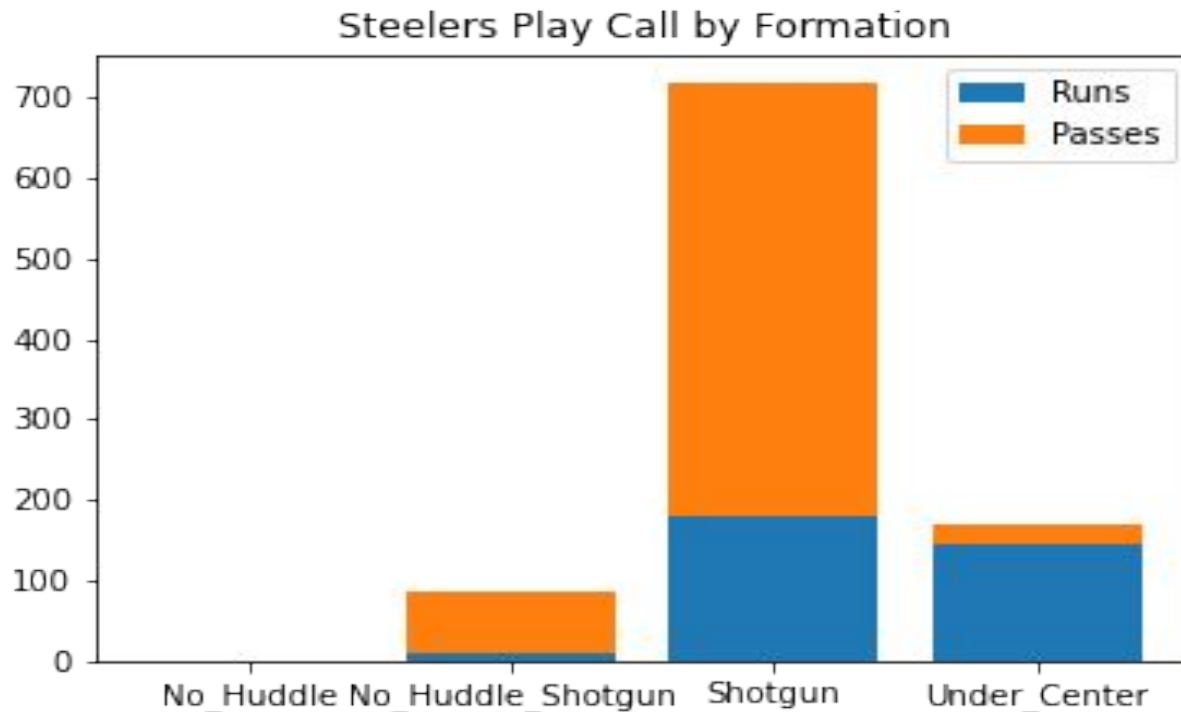
Data

- Downloaded a csv file from <http://nflsavant.com/about.php> containing all 46,289 nfl plays from the 2020 season
- Features include
 - General info
 - the defensive team, date of game
 - Pre-snap info
 - offensive formation, down, distance, etc.
 - Post-snap info
 - Yards gained, whether or not the play resulted in a touchdown
- Added run/pass defensive ranks to the data from <https://www.pro-football-reference.com/years/2020/opp.htm>
 - Run rank criteria: Y/A, Pass rank criteria: NY/A

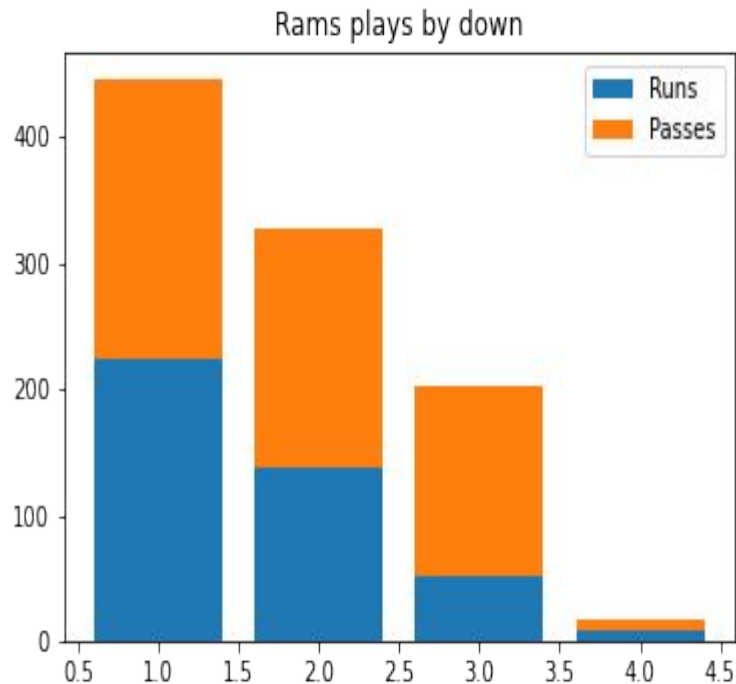
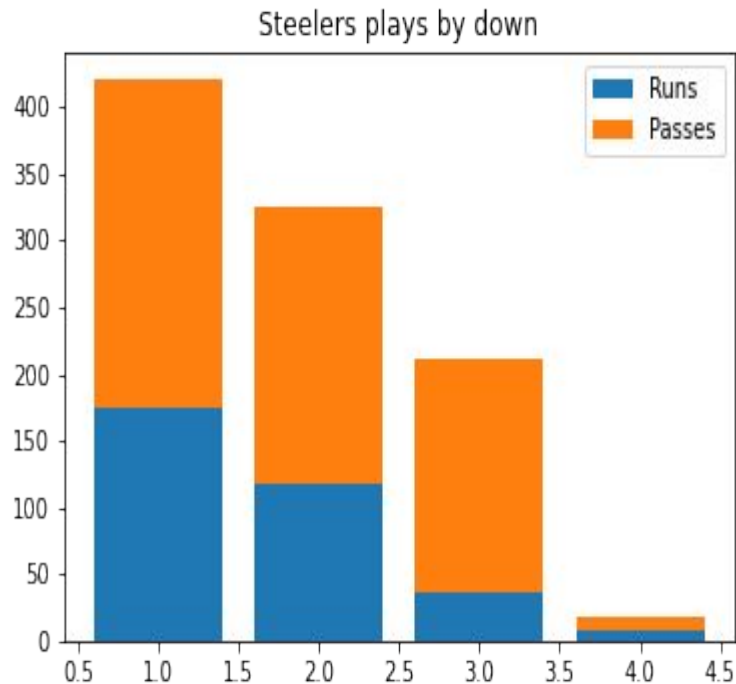
Pre-Processing

- Removed plays that weren't either a run or pass
 - punts , field goals, extra points
 - Ended up with 29,390 plays
 - About 918 plays per team
- Dropped columns that related to what happened after the play
 - Yards gained, did a touchdown occur, did an interception occur etc.
- Scaled the data
- Factors used in modeling: down, distance, defensive team ranks, formation of the offense
- Converted the down and formation variables into dummy variables

Exploratory Data Analysis



Exploratory Data Analysis



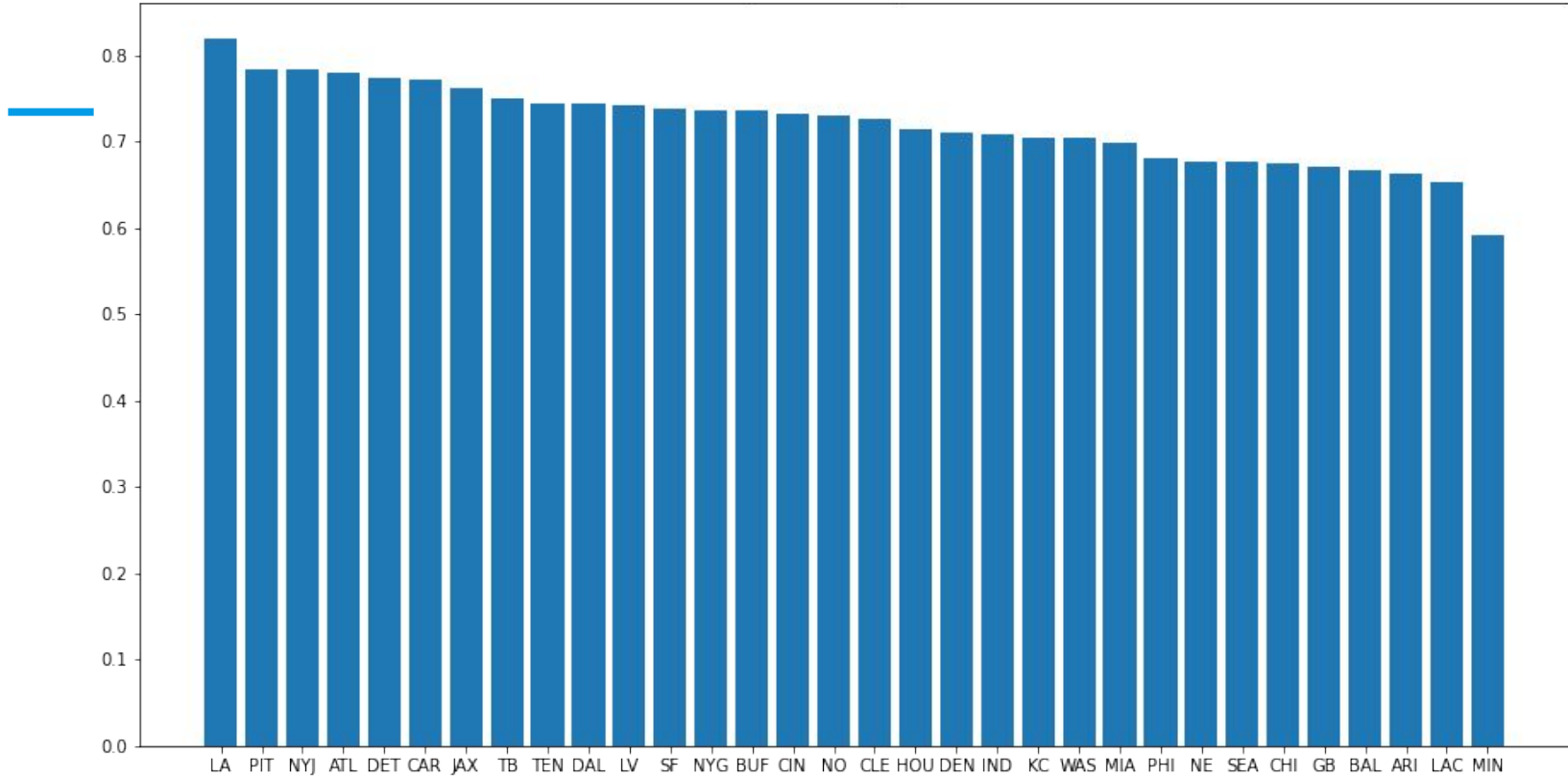
Modeling

- Baseline accuracy for all the data was 0.589554%
 - If the model was simply guessing a pass every time it'd be right 58.95% of the time
- Filtered through all 32 teams and ran several classification models
 - Logistic Regression, KNearestNeighbors, Decision Tree, Bagged Decision Tree, Random Forest, AdaBoost, Support Vector classifiers
 - Used a grid search to find the best hyperparameters
 - Neural Network with TensorFlow
- Analyzed the Logistic Regression Classifier to look at the impact of each predictor on each team
 - Compared the differences among the teams

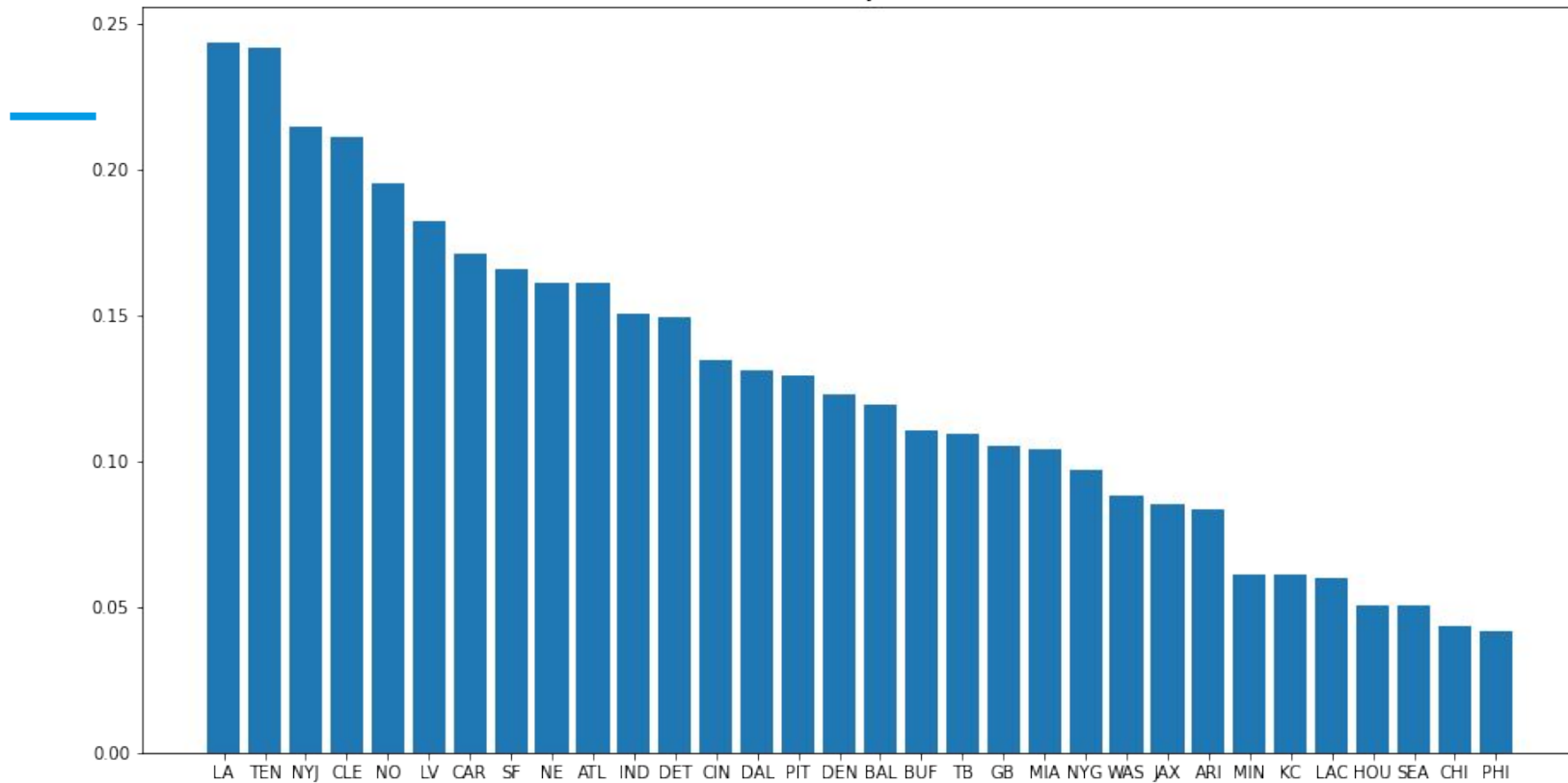
Checking the accuracy of the best models

- Baseline accuracy - 0.5895
- Logistic Regression average accuracy - .72
- KNN average accuracy - .68
- Decision tree average accuracy - .66
- Bagged Decision Trees average accuracy - .68
- Random Forest average accuracy - .69
- AdaBoost average accuracy - .71
- Support Vectorizer average accuracy - .72
- Neural Net average accuracy - .72

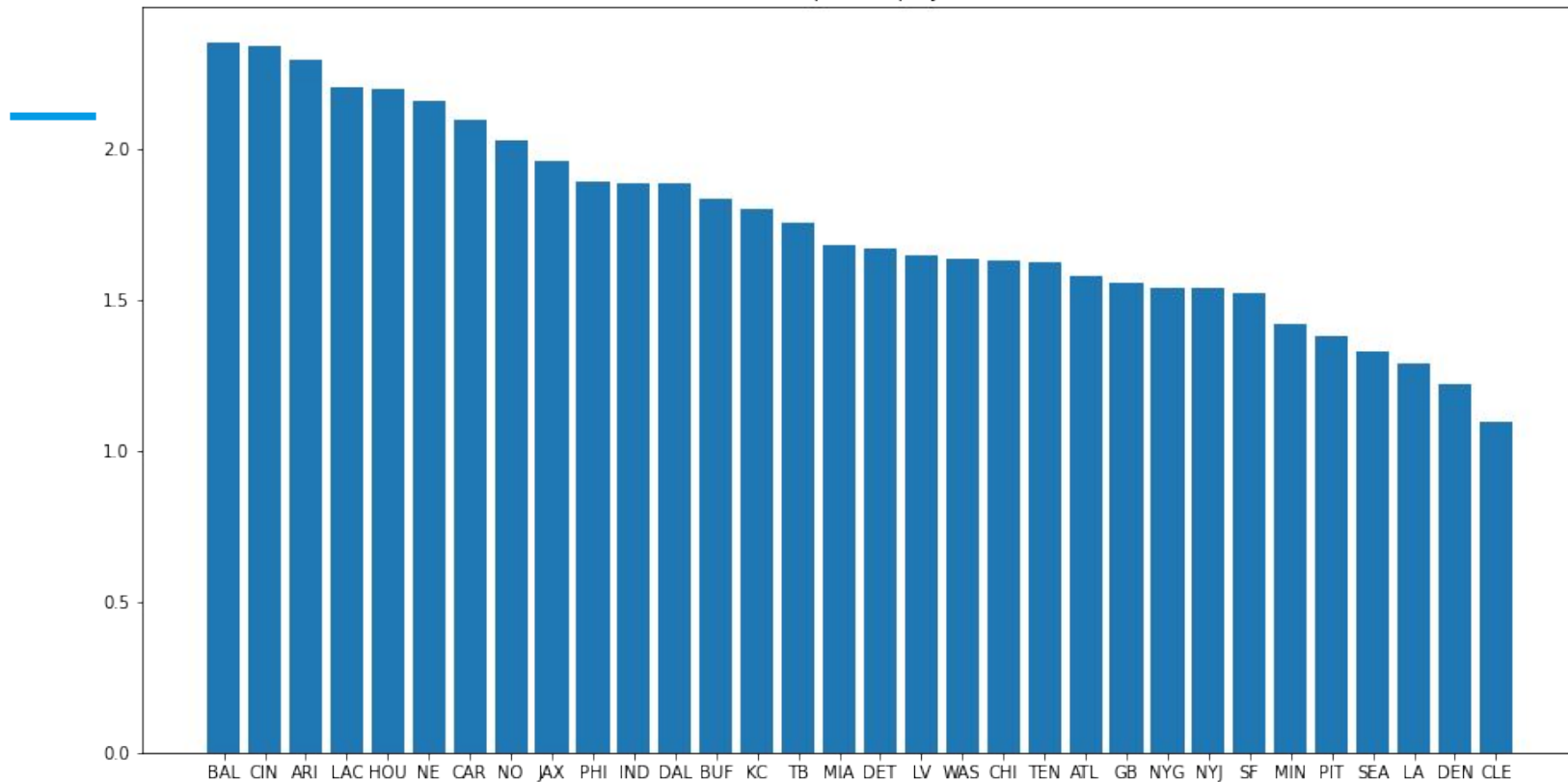
Accuracy Scores for Log Classification



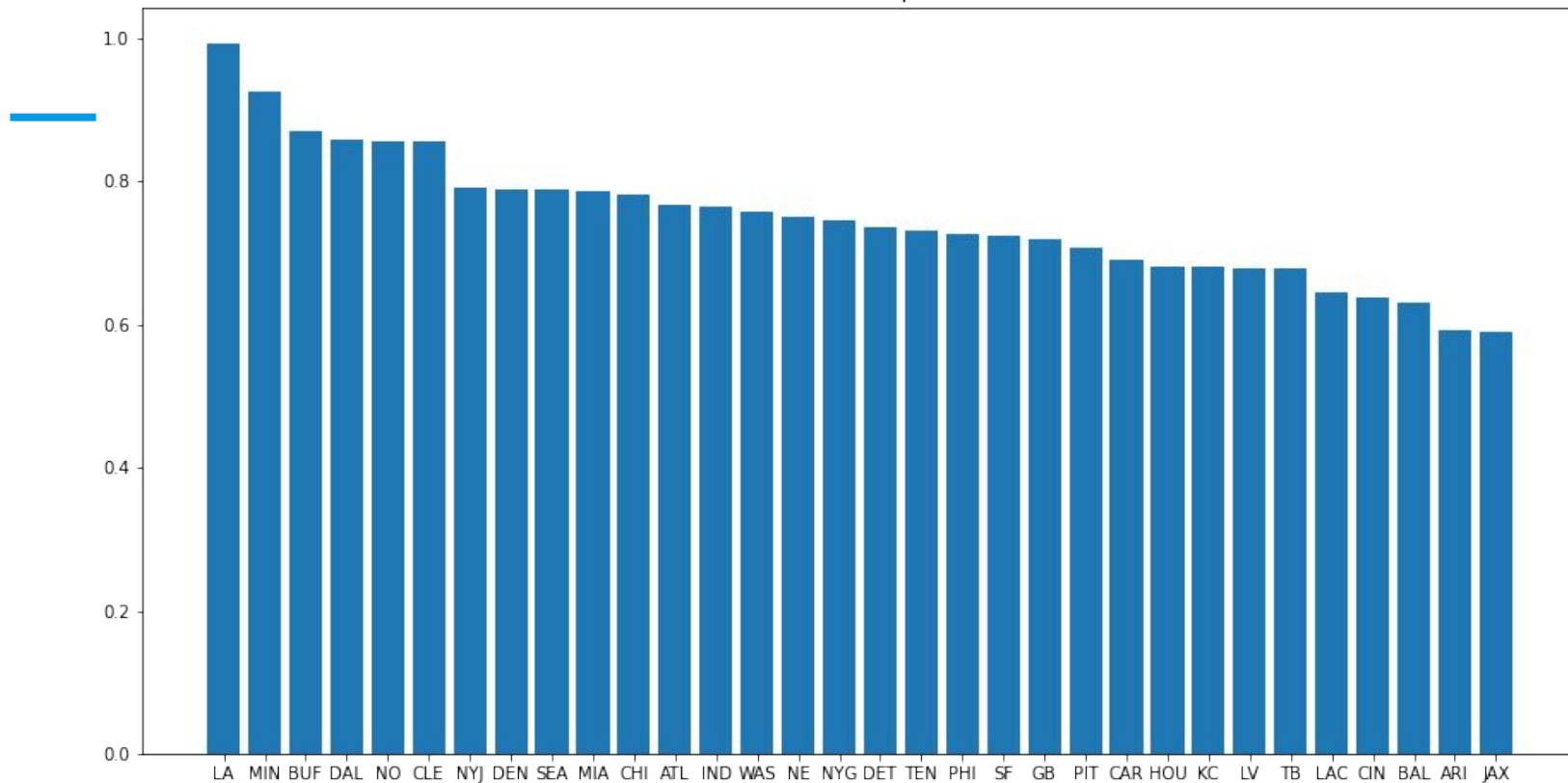
Model accuracy vs Baseline



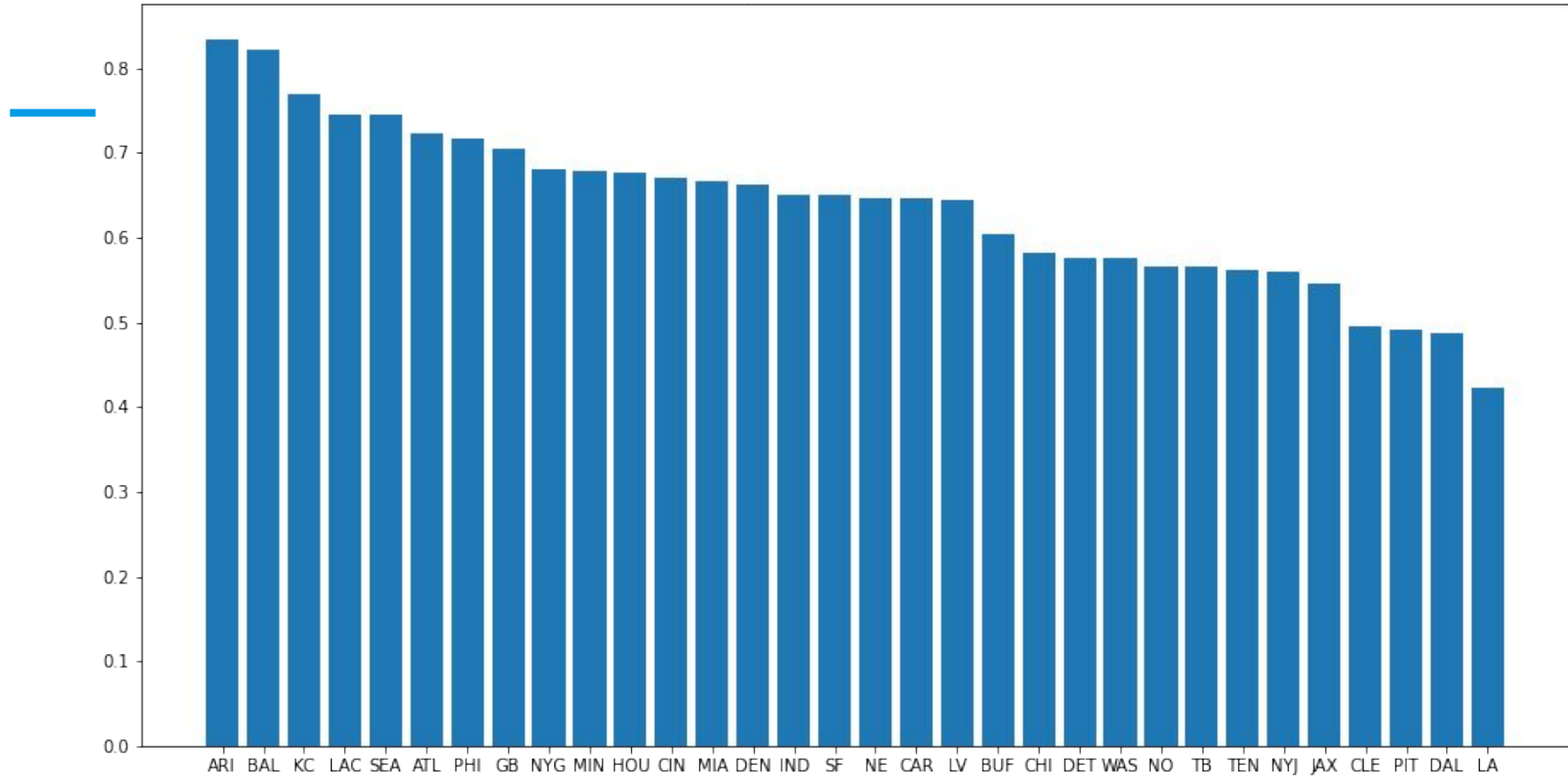
Yard impact on play call



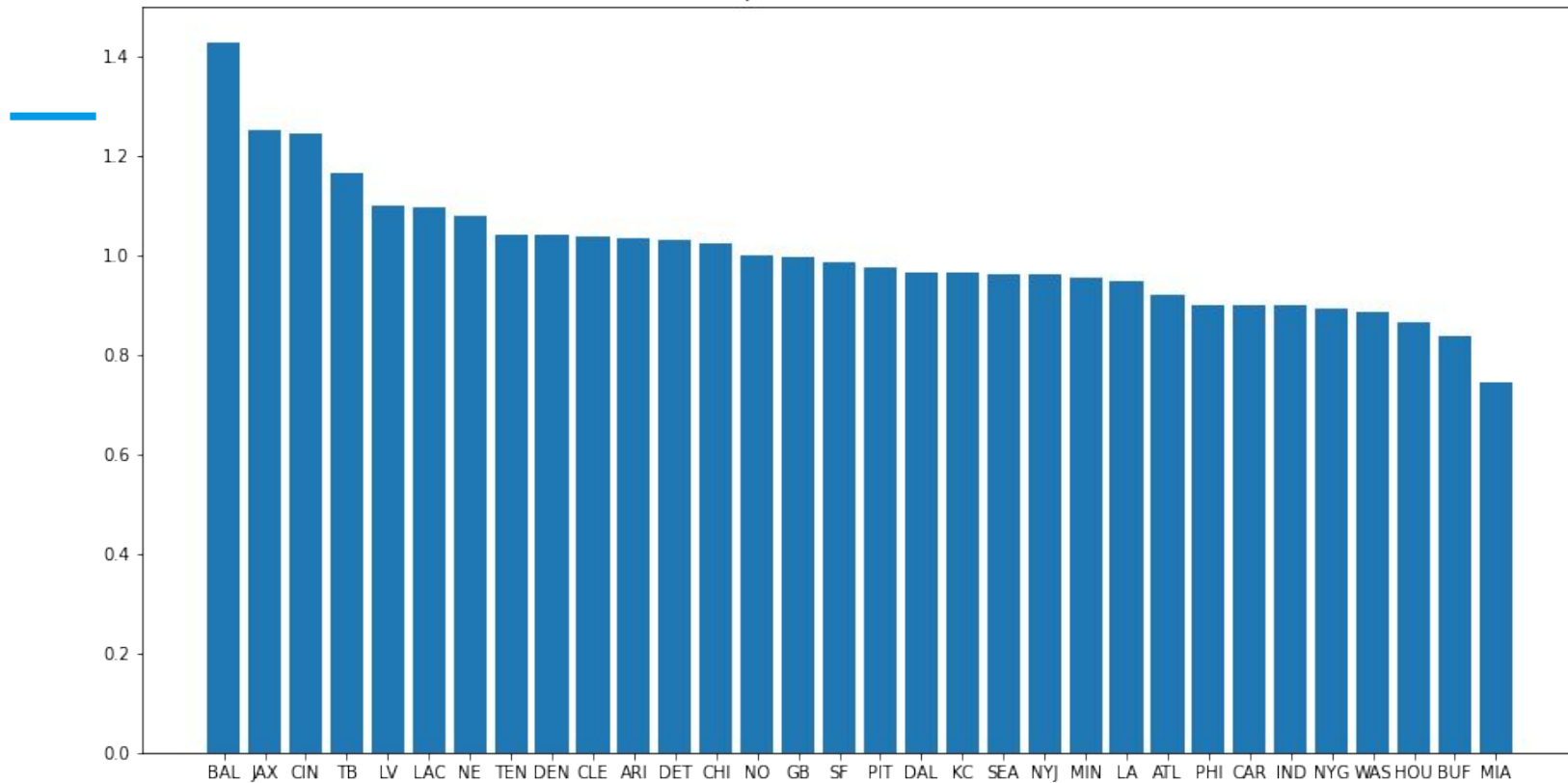
1st Down Impact



Impact when under Center



Impact of the Pass Defense



Conclusions

- Most predictable teams: Rams, Steelers, Jets
- Distance
 - The Browns' play calls were the least impacted by distance
 - Run less when they're close, pass less when they're far
- Down
 - Rams' play calls were by far the least impacted by down
 - The down wasn't as helpful in predicting the Rams' play calls
- Opponent
 - Ravens exploited their opponents' weaknesses the best
 - Ran against bad run defenses, passed against pass defenses
- Formation
 - Rams' play calls were the most impacted by formation
 - It was easier to predict whether or not they would pass based off what formation they were in

For the future

- Time series modeling
 - Throughout the season
 - Take injuries into account
 - Throughout the years for a specific coach
- Deep dive into 4th downs and 2 point conversions
 - Plays are more specific than a general play
 - Easier to learn information about tendencies
- Include better information on the offensive and defensive formations
- Take game score into account



The End

