Understanding Yards in the National Football League

Sharon Kwak

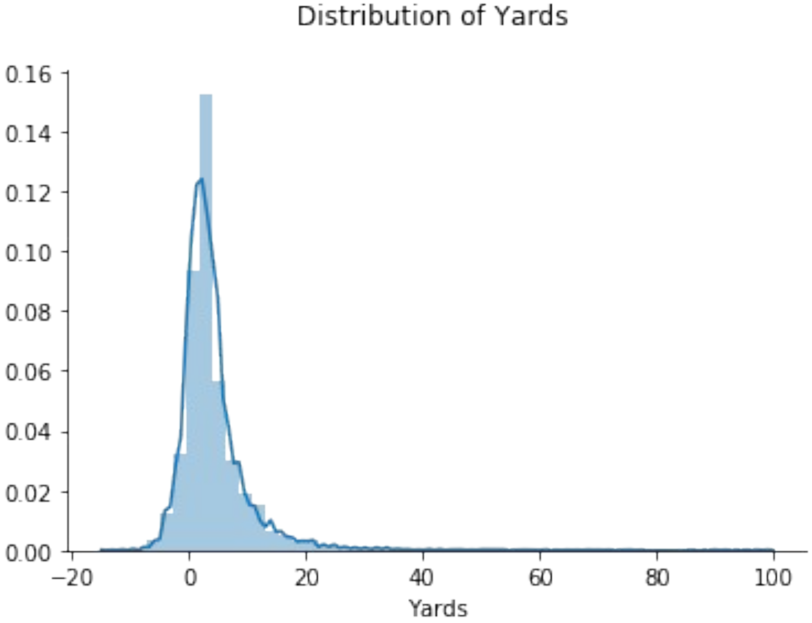
Predicting Yard Gain

- Give more insight to coaches in deciding better plays based on various conditions of the game
- Help fans predict more accurately the results of each handoff

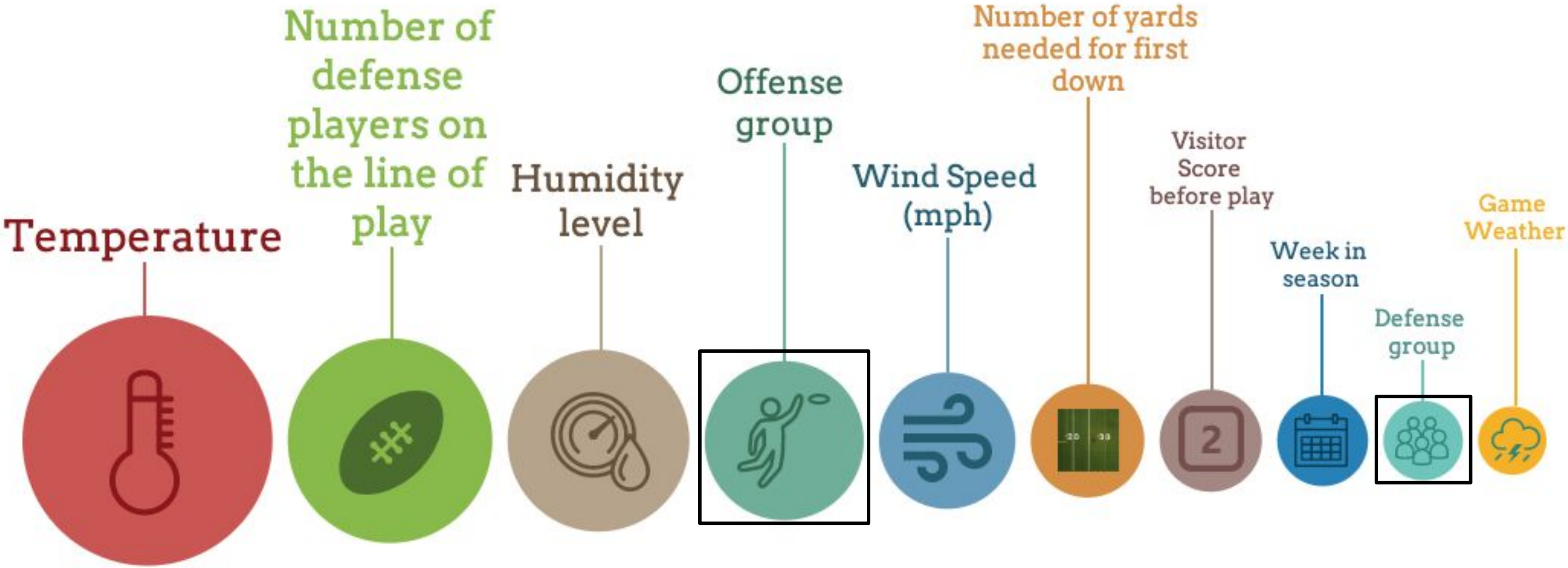
Data

- Kaggle
- Play-by-play data for each handoff (Quarterback hands ball to Running Back)
- 37 features include:
  - player (name, height, weight)
  - team (opponent, home/away)
  - stadium (indoor/outdoor, grass type)
  - weather (temperature, wind, humidity)

Which factors of each play most impact the number of yards gained?



Top 10 important features in predicting yards



Temperature

Number of defense players on the line of play

Humidity level

Offense group

Wind Speed (mph)

Number of yards needed for first down

Visitor Score before play

Week in season

Defense group

Game Weather

Example Predictions

Coaches can decide better Offense and Defense groups based on the top factors (e.g. temperature)

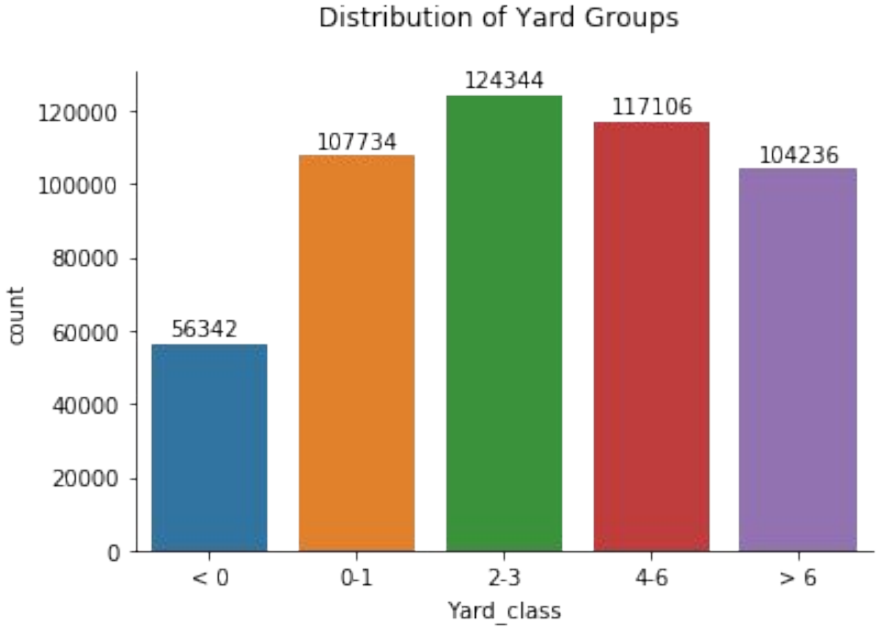
- Low temperature, need more yards
  - 7 Offense Linemen
  - 1 Running Back
  - 2 Wide Receivers

- Low temperature, want less yards for opponent
  - 3 Defense Linemen
  - 4 Linebackers
  - 4 Defensive Backs (Safety, Cornerback)

Next Steps

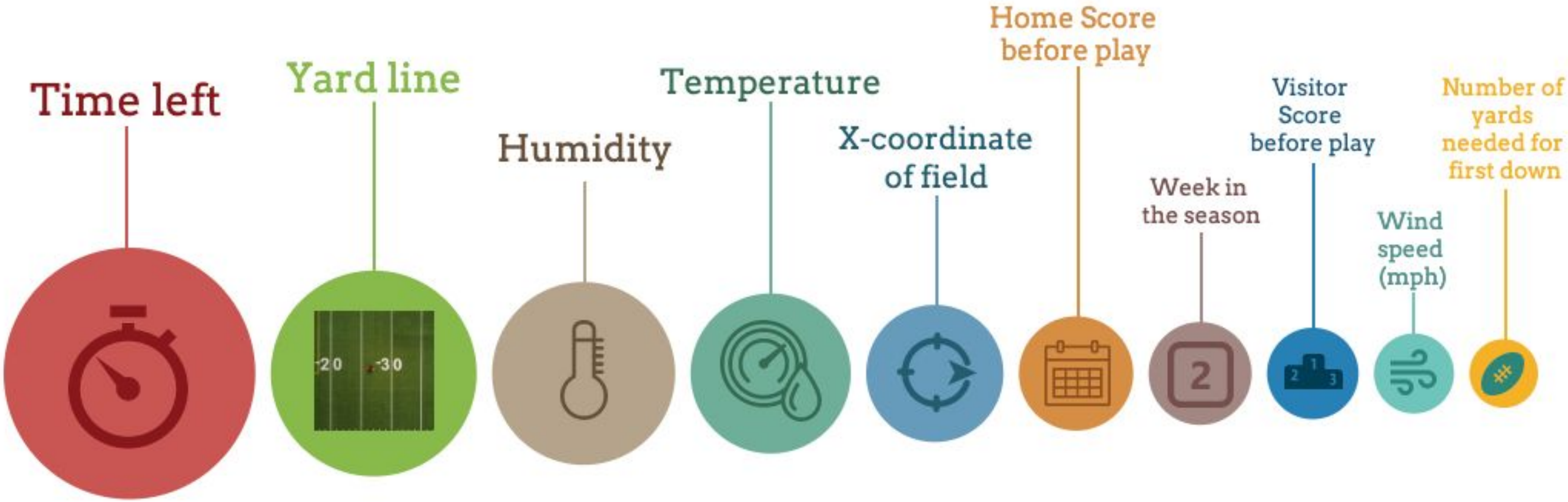
- Include player statistics from http://www.nfl.com/stats/player
- Include statistics from other types of plays (throwing or passing)
- Run deep learning time-series RNN (Recurrent Neural Network)

Which factors of each play most impact the range of yards gained?



- Yards were binned based on the five-point summary (box plot)
  - < 0: Negative yards (includes the minimum)
  - 1: 25th quartile
  - 3: Median (50th quartile)
  - 6: 75th quartile
  - > 6: Rest of the data (includes the maximum)

Top 10 important features in predicting range of yards



Time left

Yard line

Humidity

Temperature

X-coordinate of field

Home Score before play

Week in the season

Visitor Score before play

Wind speed (mph)

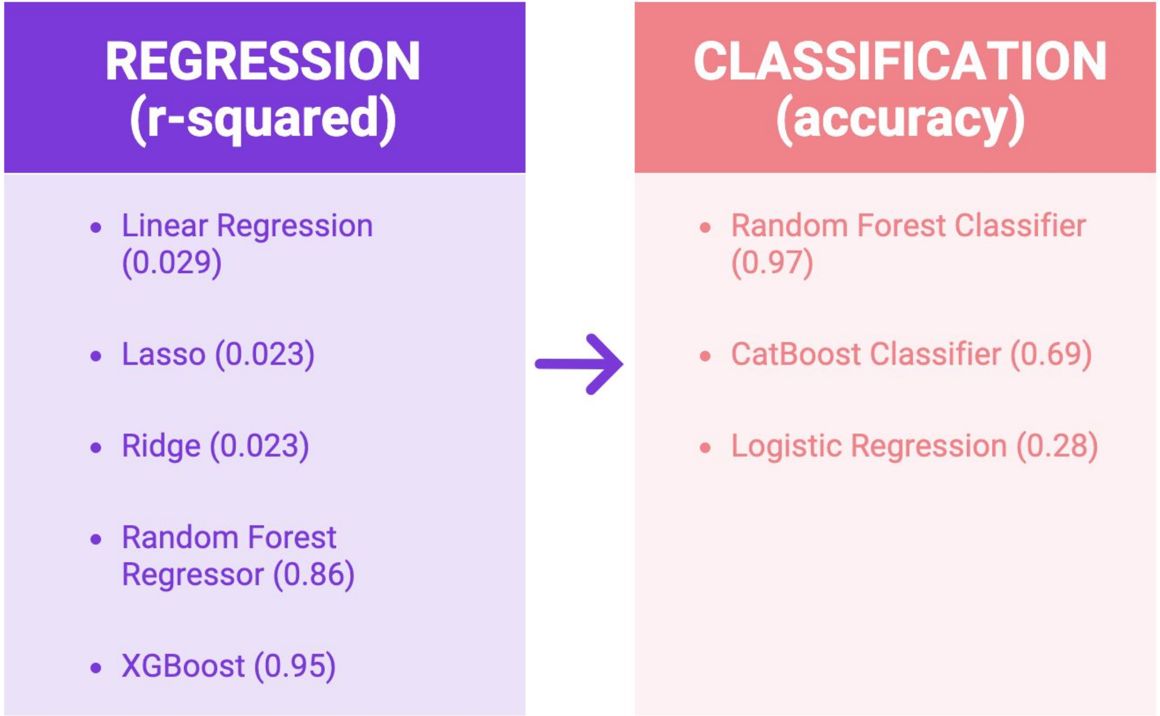
Number of yards needed for first down

Example Predictions

Coaches and players can know for which conditions more practice is needed to run more yards

- Less than 4 minutes left in quarter
- 50+ yards left to touchdown
- Around 40-50 degrees
- Low wind speed (around 6 mph)

Model Results



REGRESSION (r-squared)

- Linear Regression (0.029)
- Lasso (0.023)
- Ridge (0.023)
- Random Forest Regressor (0.86)
- XGBoost (0.95)

CLASSIFICATION (accuracy)

- Random Forest Classifier (0.97)
- CatBoost Classifier (0.69)
- Logistic Regression (0.28)