



# Classifying NFL Plays



Angela Huang



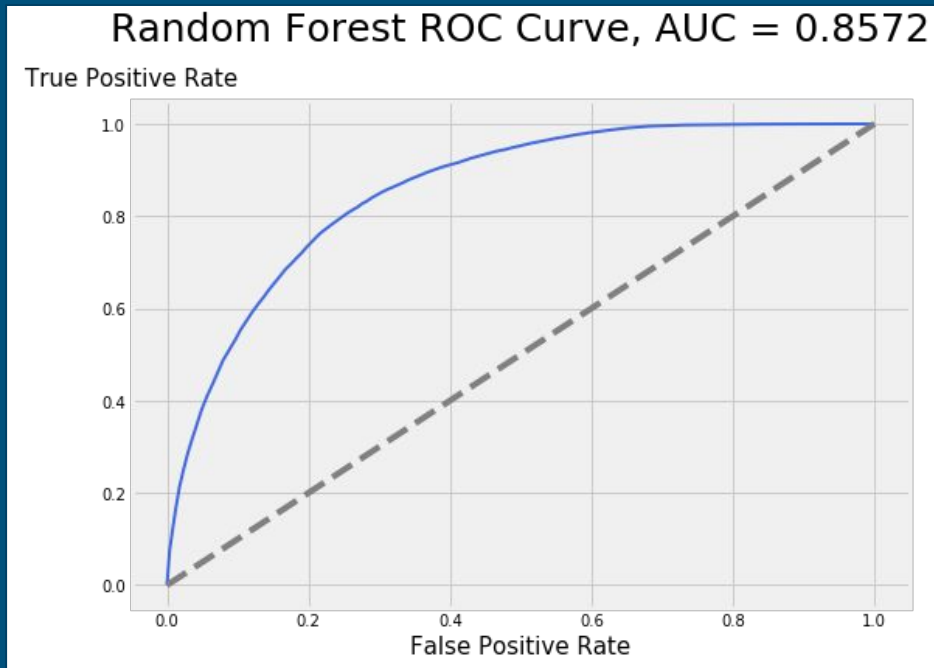
# Background

- Adoption of data analytics in football
- Objective:  
Will a play be positive yards or not?



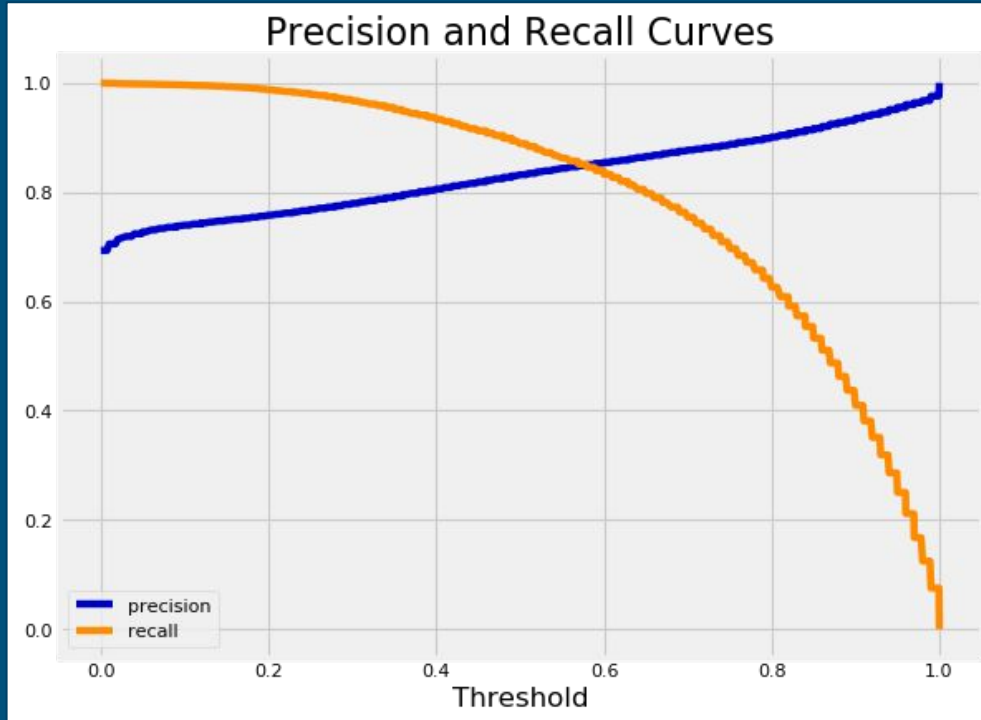
# Data & Methodology

- Play by play data from 2009 through 2018 (week 15)
- PostgreSQL → pandas
- Models tested:
  - Logistic Regression
  - KNN
  - Naive Bayes - Gaussian
  - SVC
  - Decision Trees
  - Extra Trees
  - **Random Forest**



# Results

- Feature importance
  - Improved AUC by 3.74%
- Adj threshold for F1 score
  - Threshold: 0.39
  - F1 score: 0.867



# Tableau Dashboard Demo

---

# Future Work

---

- Additional features from other sources
- Increase Tableau dashboard functionality
- Optimize for teams & coaches

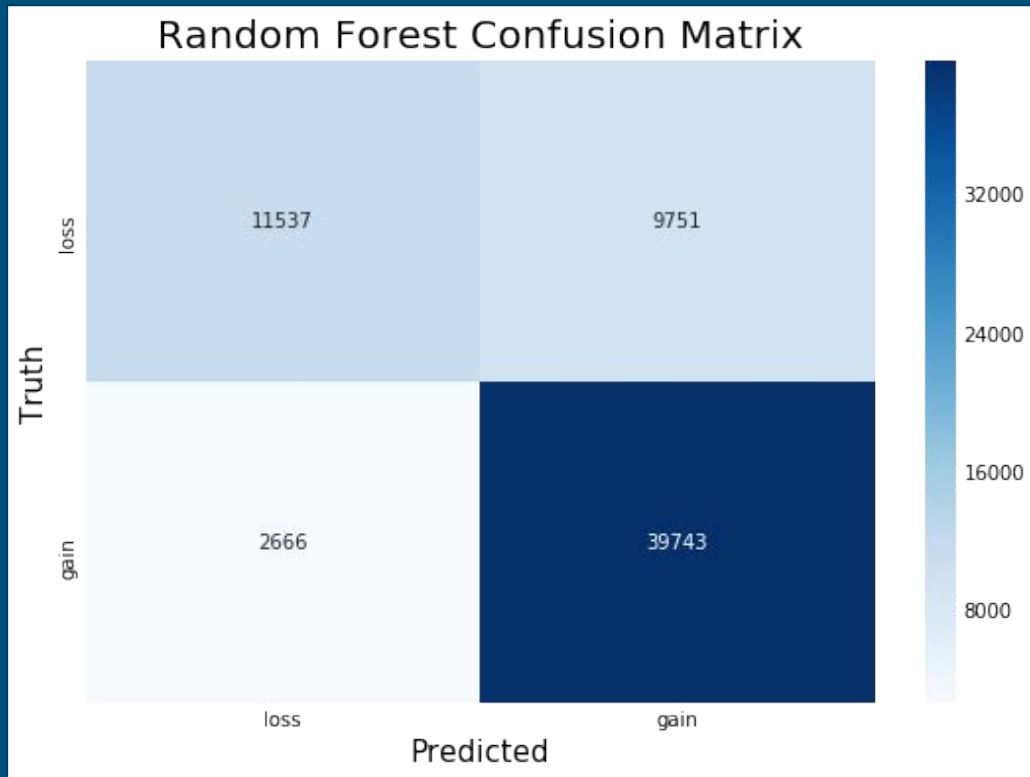


# Thank You!

---

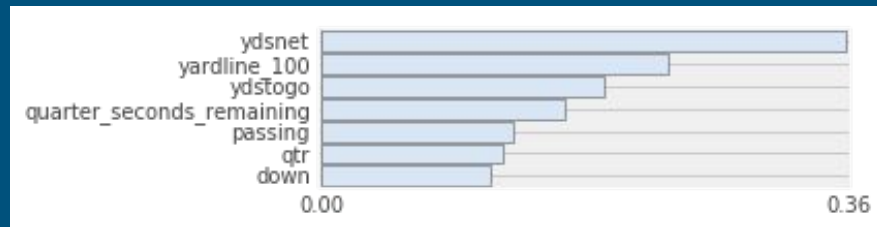
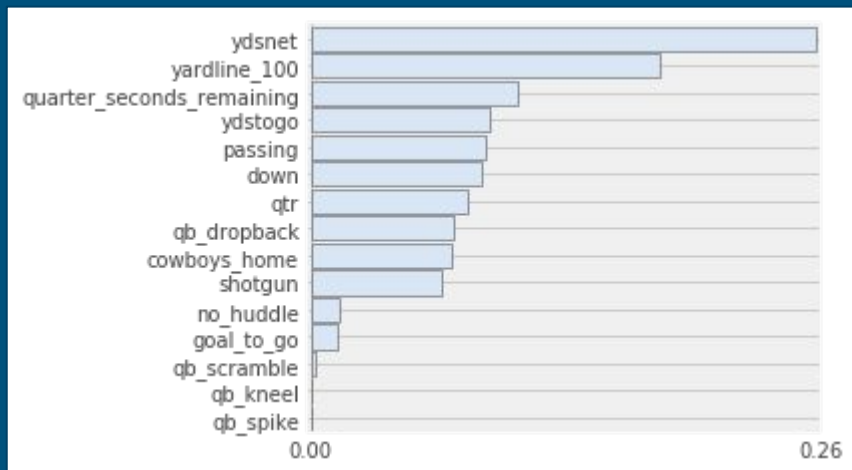


# Appendix - Random Forest Confusion Matrix

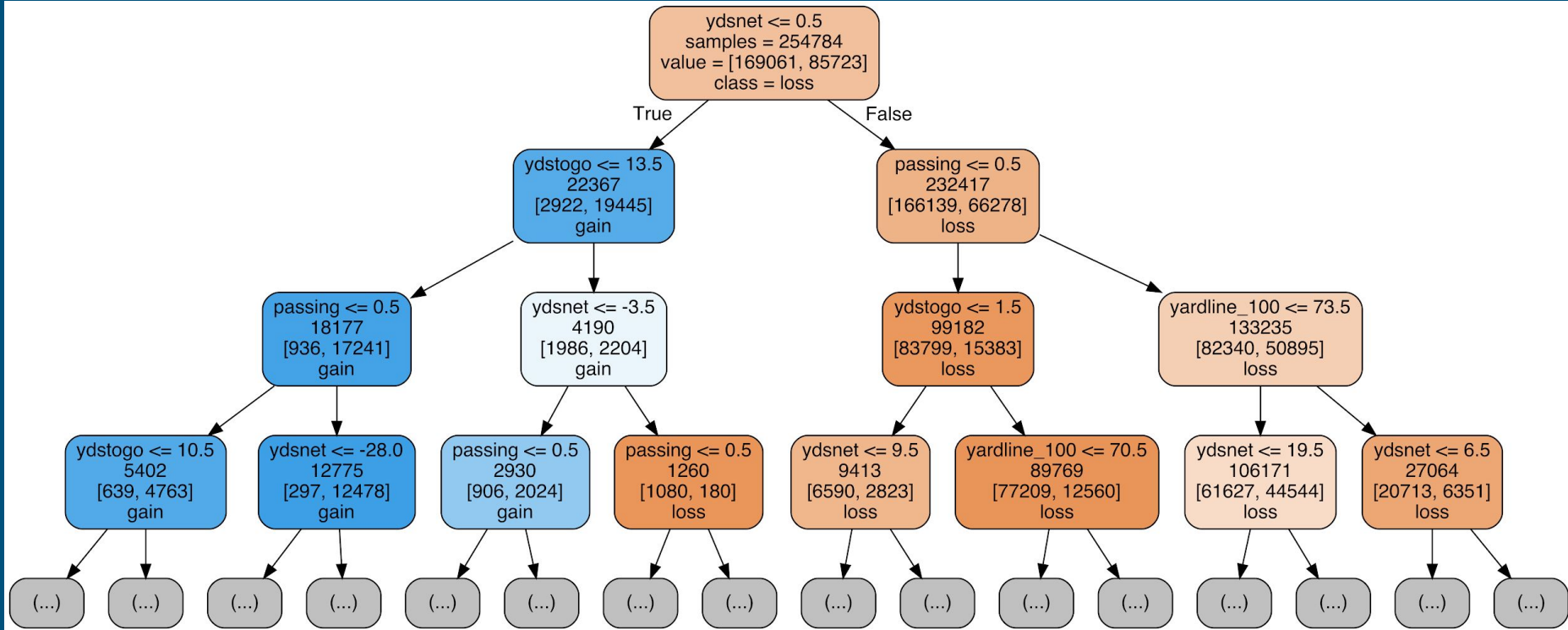




# Appendix - Important Features



# Appendix - Sample Decision Tree



# Appendix - Tools Used

---

- PostgreSQL
- Pandas
- Numpy
- Matplotlib
- Seaborn
- Scikit Learn
- rfpimp
- TabPy
- Tableau