



# Predicting NFL Team Scores

Greg Niss

## Secular Industry Growth for Sports Gambling




*State lawmakers around the country may look to legalize sports betting as a way to raise revenue. Commercial casinos generated over \$10 billion in state and local gaming tax revenue last year, and that doesn't include sales and income tax tied to casino gaming. State governments have gotten addicted to the lucrative tax revenue from casinos; legalizing sports betting is the next logical way to get a little more juice out of the same orange. - Forbes Magazine July 1, 2020*

## Secular Industry Growth for Sports Gambling



- *Three states had measures on the ballot to legalize sports betting: Maryland, Louisiana and South Dakota. All voted yes.*
- *Overall, there are 19 states as well as Washington D.C. where sports betting services are currently operational, according to the American Gaming Association.*
- *Six other states, including the three that passed the measure in the election, have authorized legal sports betting but have yet to launch it.*

## Data vs Goliath: Is it Possible to match the oddsmakers' accuracy at predicting Competitive Sports?



*“Have you ever listened to an oddsmaker? They know the subsurfaces at the various fields. They know what time the tide comes in in San Francisco because it softens the playing surface. They know the average wind speeds and temperatures in Buffalo by the day.”*

~ Quote from Stanford Wong's, Sharp Sports Betting

# Objective:



As someone who is optimistic about participating in the long term prospects of the sports gambling industry, an essential first step in understanding the industry better is to examine the factors that are most important to oddsmakers in determining point spreads and over/under lines for NFL games.

---

# Outline

1

Data

2

Process

3

EDA

4

Modeling

5

Evaluate

6

Conclusions and Next Steps

# Data

- **NFL Seasons 2010 – 2016** - Two sources of data truncated
- **National Football League (NFL) game results since 1966 with betting odds information** since 1979. Dataset was created from a variety of sources including games, scores, betting lines, stadium data and weather from a variety of public websites such as ESPN, NFL.com, and Pro Football Reference with weather information from NOAA data and NFLweather.com.
- **Individual player statistics database going back to the 1940s.** Over 25,000 players who have played over 1,000,000 football games.





# Process: Features

---

- **Moving Average Score** - How many points did the target team and opponent team score in the 7 weeks prior to this contest.
- **Moving Average Margin of Victory**
- **Moving Average Yards** - How many passing and rushing yards did the target team and opponent team gain in the 7 weeks prior to this contest.
- **Home Team** - Playing at home has typically provided a 2.5 to 3 point advantage to the home team.





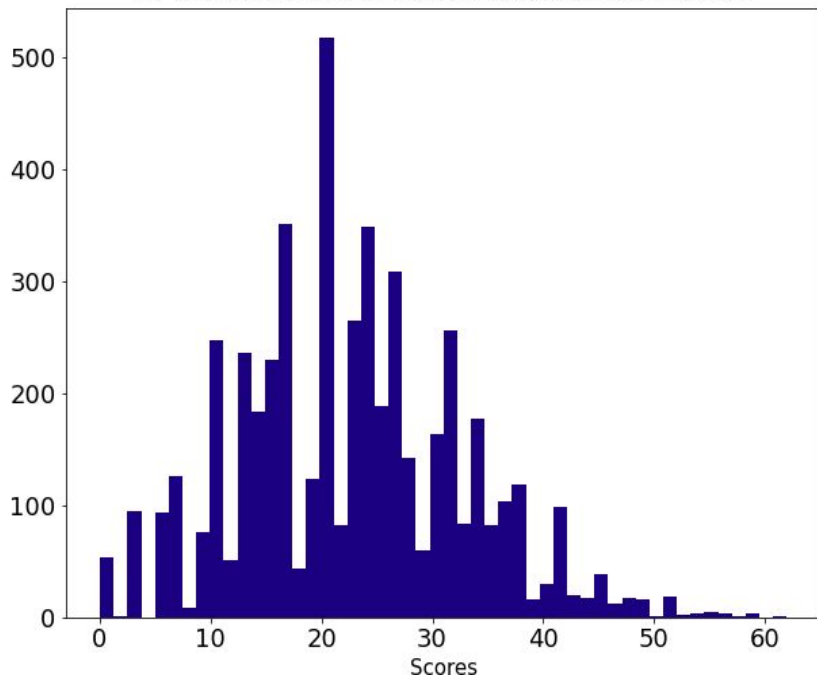
# Process

- Take weekly snapshot of target team's statistics
- Calculate rolling average of features
- Map all opposing team's statistics to target team
- Run Models on data in aggregate, not in time series(yet)
- Significant challenges: mapping and filtering down data set; usually a classification prediction

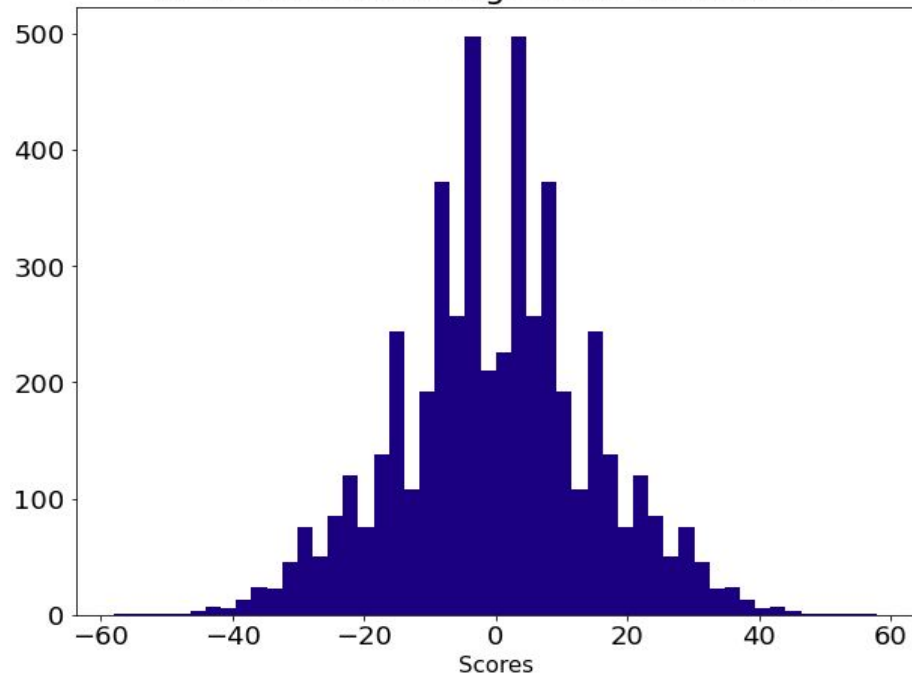


# EDA: Score Distributions

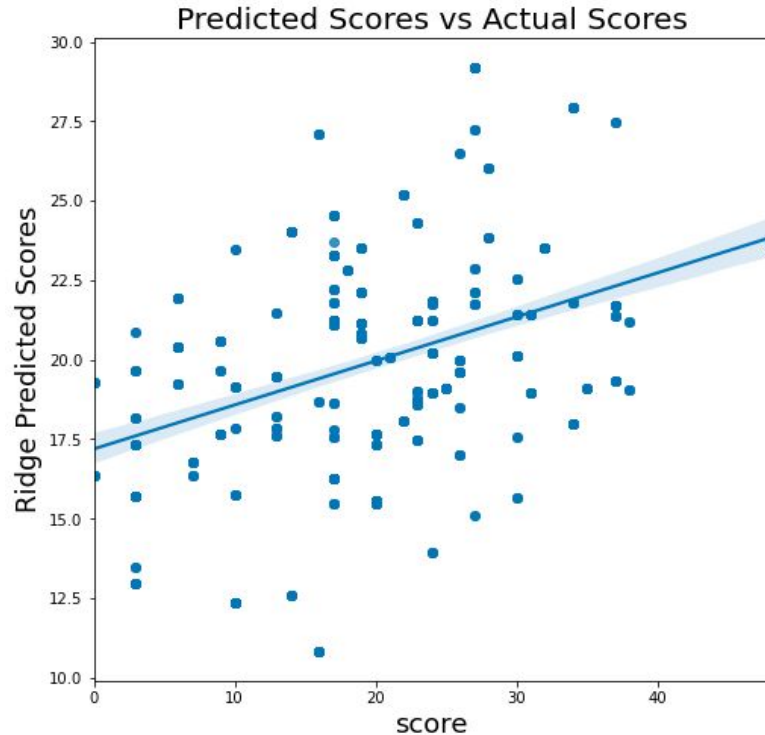
NFL Scores Distribution from 2010 - 2016



NFL Scores Win Margin from 2010 - 2016



# EDA: Artificially High Predicted Scores



# Model Evaluation

**Baseline:** (Using 7 week moving average)

MAE: 8.05

**Linear:**

MAE: 7.38

Train Score: 0.126

Test Score: 0.153

**RidgeCV:**

MAE: 7.38

Train Score: 0.125

Test Score: 0.152

**AdaBoost:**

MAE: 5.28

Train Score: 0.628

Test Score: 0.628

**GradientBoost:**

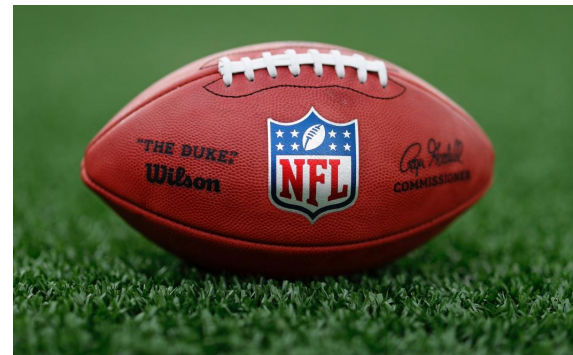
Train Score: 0.999

Test Score: 0.999

**Sequential:**

MAE: 3.00

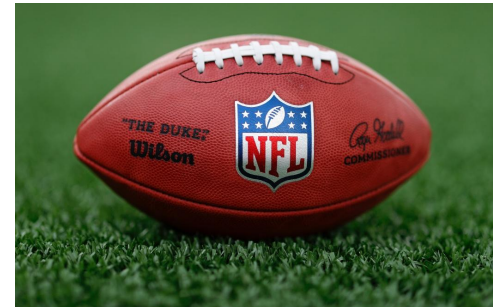
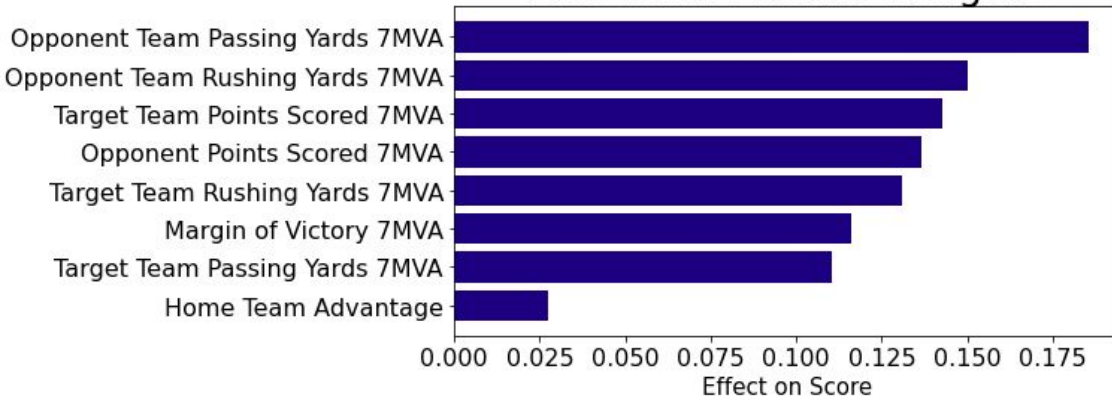
MSE: 5.00



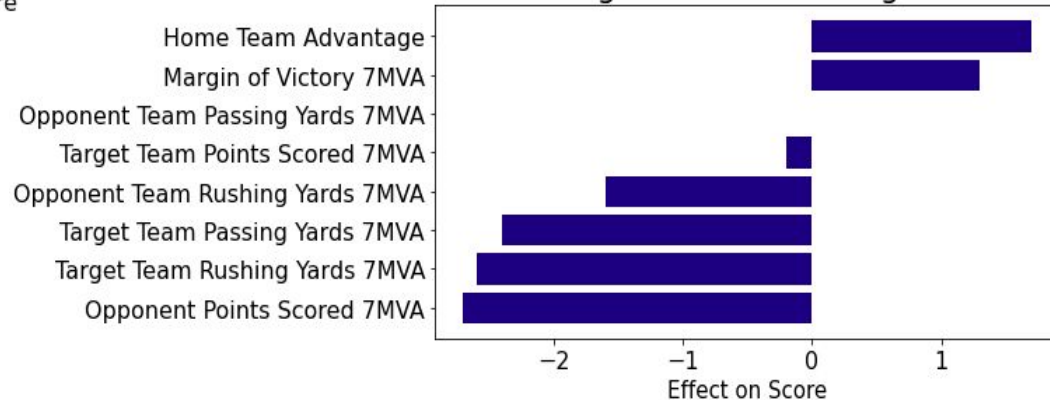


# Evaluation: Weights and Autocorrelation

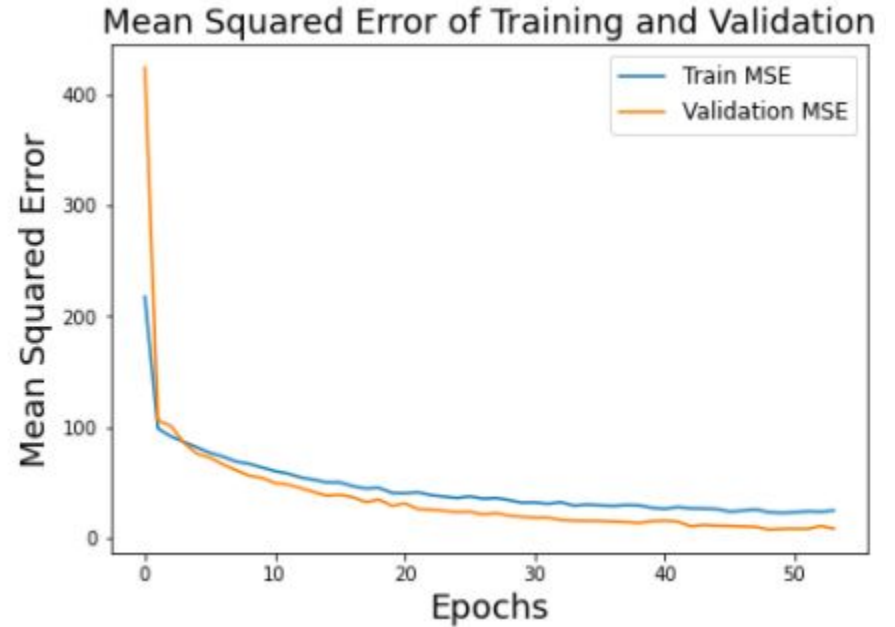
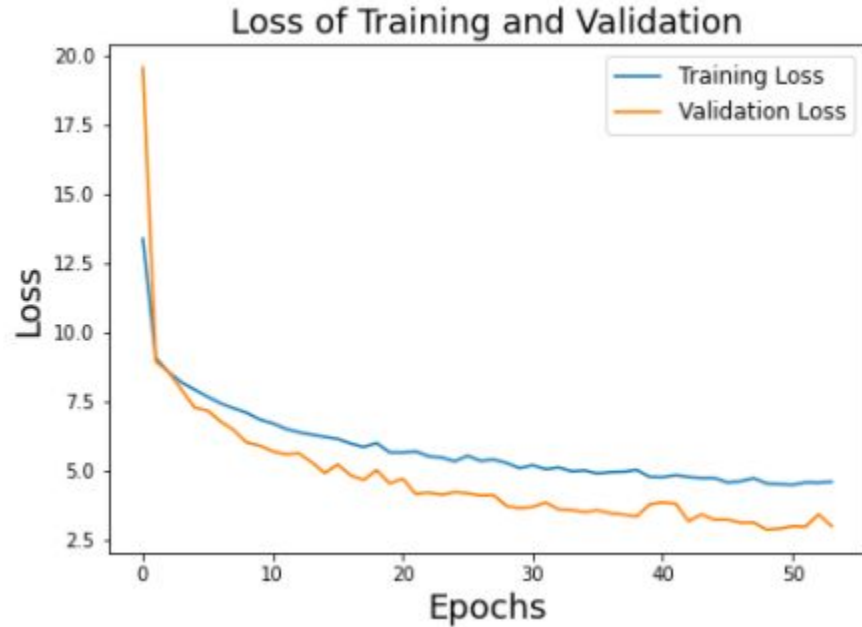
## AdaBoost Feature Weights



## Ridge CV Feature Weights



# Models: Sequential





# Next Steps:

---

- Add more data so can get more granular. Individual player stats, weather -bigger factor.
- Enhancing and working with multi dimensional models time series, Bayesian, RNN.
- BackTesting!
- Max Kellerman - applying to less followed sports where there is more opportunities.



See you at this crazy new sportsbook in Vegas!....or a sportsbook nearby.

