NHL Game Predictions

Pete Schrader

Introduction

- Graduate of Michigan State University – 2009 in Political Science
- Currently Pursuing MA in International Relations at American University
- 20s were given to the Marine Corps.
- NHL career derailed by broken leg suffered at 12 years old.

NHL Game Data

- 7 seasons from 2011 – 2018
- 11,434 games
- Broken down by team and game statistics
- Average of 5.55 goals per game
- Average spread 2.05 goals
- Which statistics or features best predict game outcomes?

Greatest moment in sports history from: https://gfycat.com/gifs/search/darren+mccarty

So What?

- Sports gambling is a $150 billion industry.[1]
- Profits can be made above around 61.5% accuracy prediction.

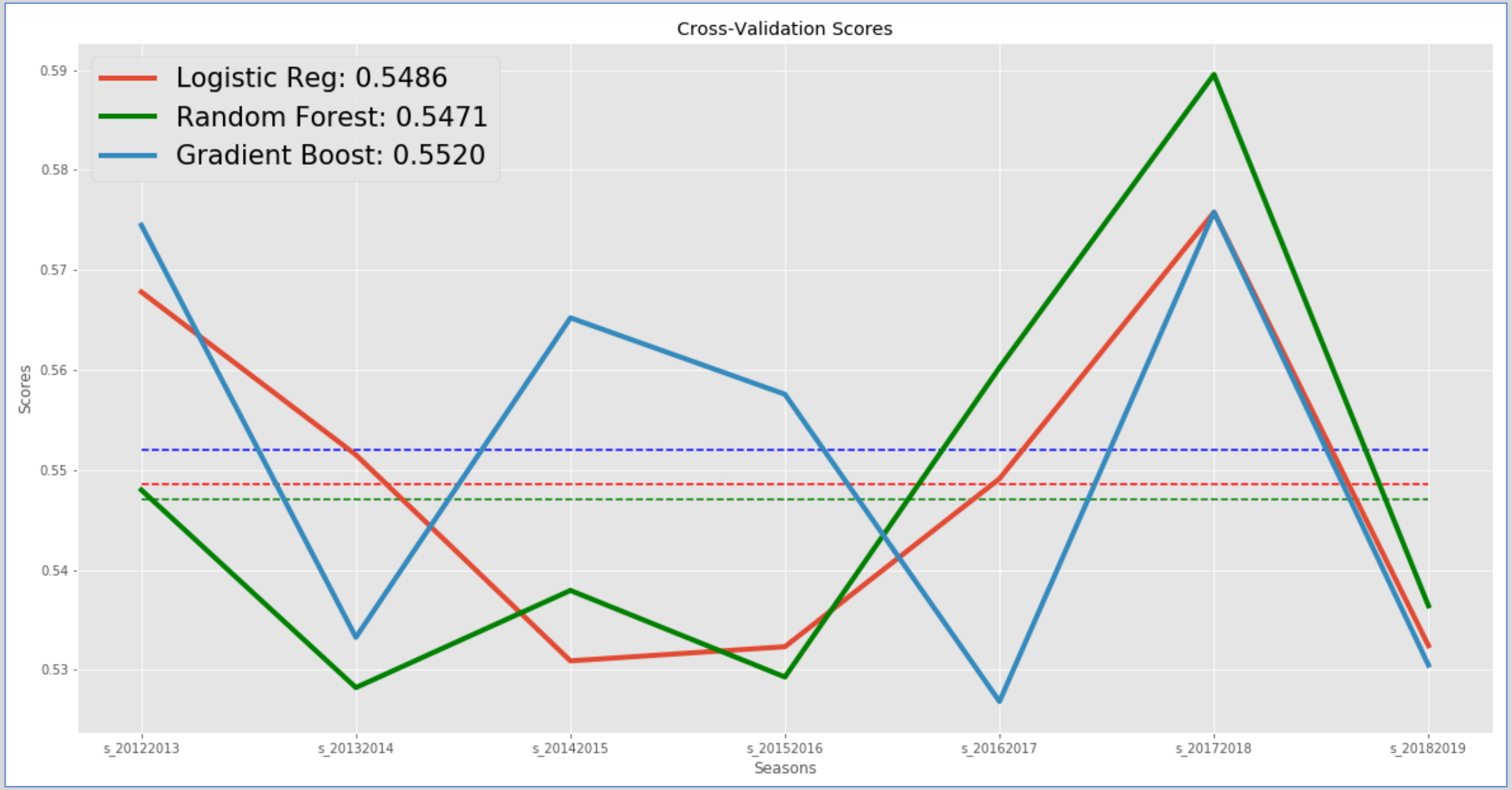
- Other Models[2]:
- Top test accuracy: 60.3% with "Logistic 3"
- Average accuracy: 59.3%
- Mentions of models performing around 62%

- *Images from corsicahockey.com*

Model Description	Validation Log-Loss	Validation Accuracy	Testing Log-Loss	Testing Accuracy
CNN 1	0.6754139	0.5784091	0.6753460	0.5900350
CNN 2	0.6742186	0.5804924	0.6762582	0.5917832
CatBoost 1	0.6725936	0.5825758	0.6727148	0.5847902
Logistic + MLP	0.6731487	0.5878788	0.6725045	0.5900350
Boosted Odds	0.6692580	0.5986880	0.6734504	0.5839161
Logistic 1	0.6736301	0.5789773	0.6753663	0.5821678
Logistic 2	0.6734568	0.5816288	0.6727997	0.5865385
Logistic 3	0.6705773	0.5916667	0.6703581	0.6031469
Logistic 4	0.6705278	0.5882576	0.6683209	0.5970280
Logistic 5	0.6769002	0.5732955	0.6736581	0.5900350
Logistic 6	0.6783505	0.5721591	0.6790032	0.5882867
Logistic 7	0.6740068	0.5767045	0.6750339	0.5882867
Logistic 8	0.6794484	0.5638258	0.6835999	0.5743007
Implied Odds	0.6712447	0.5893939	0.6700579	0.5952797
Salad	0.6675118	0.5886364	0.6712618	0.5847902
Average	0.6710315	0.5915417	0.6716220	0.5926573
Ensemble	0.6687927	0.5962500	0.6703090	0.5900350

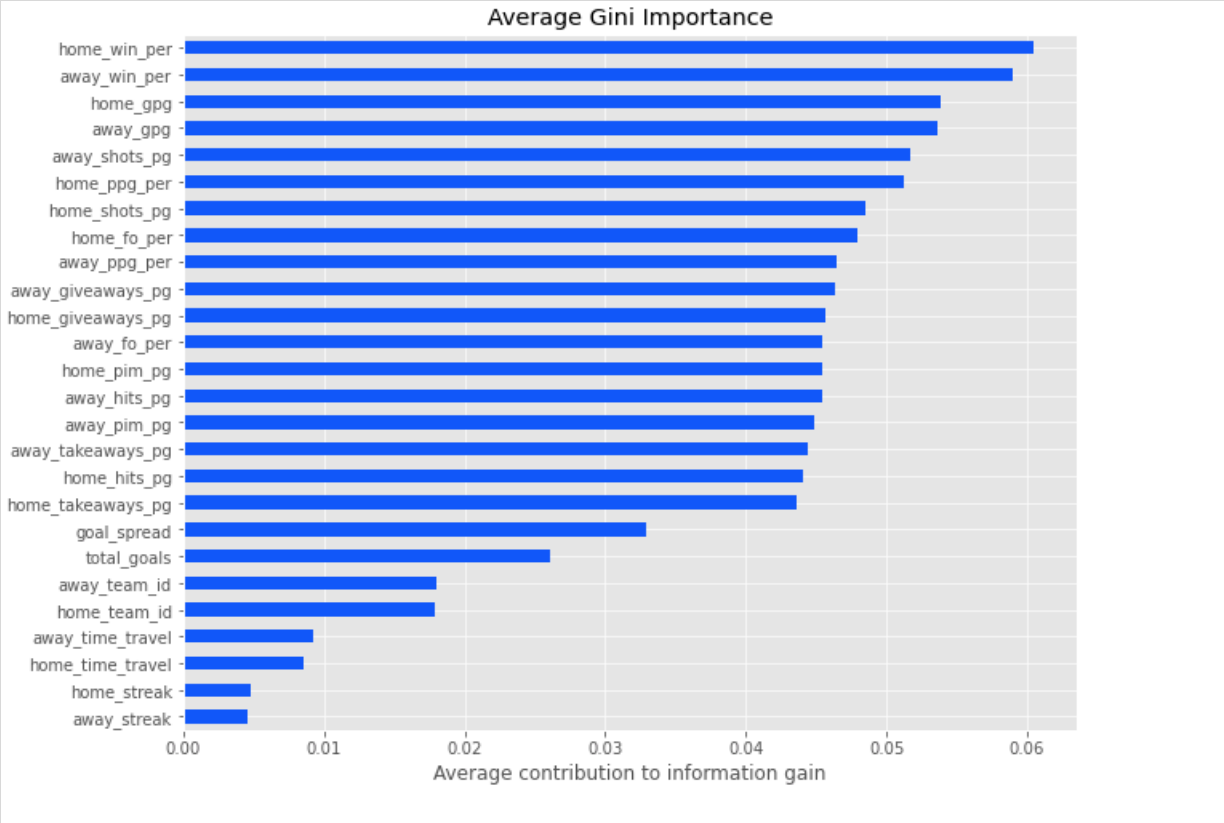
Models - All Seasons

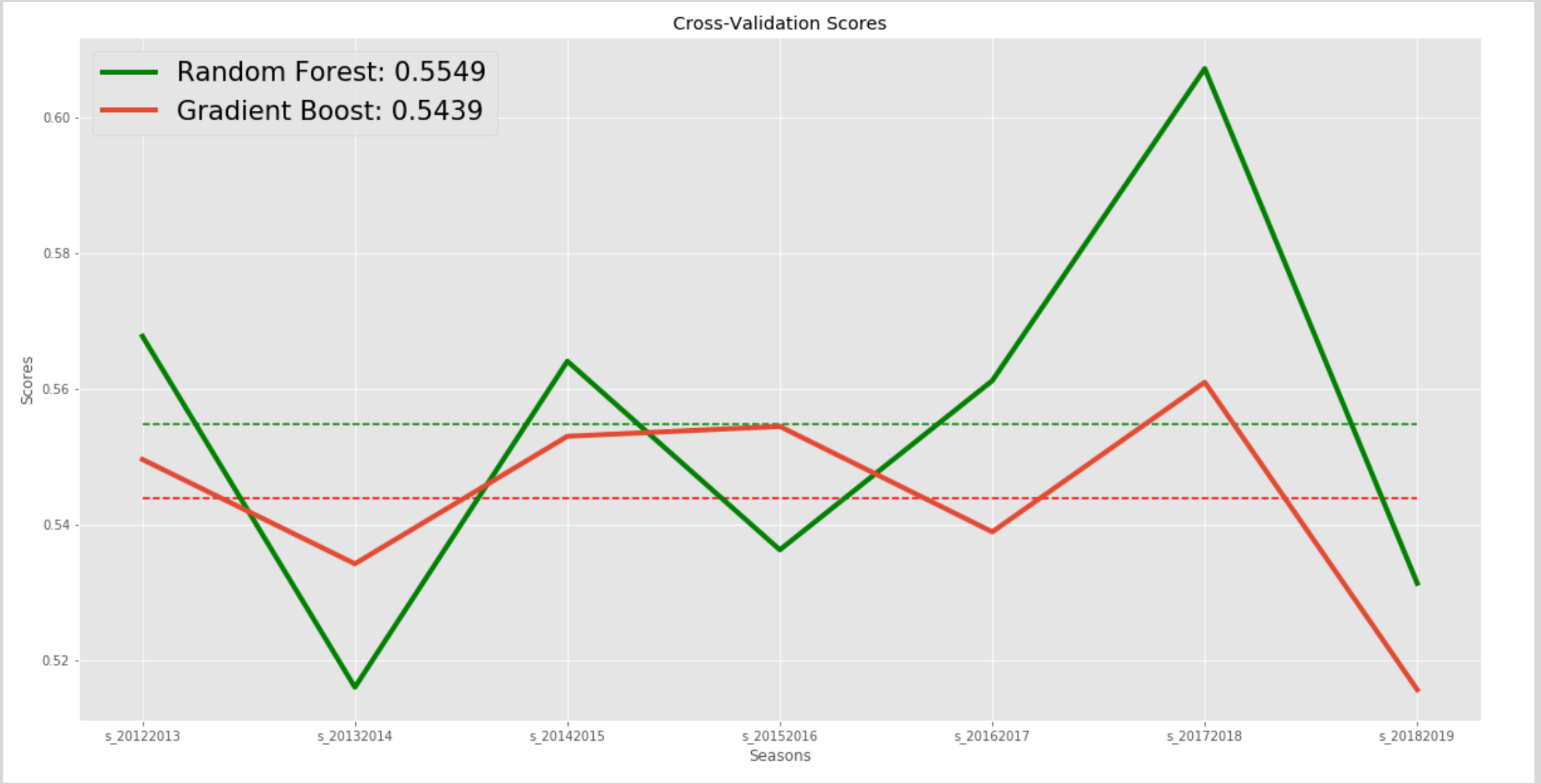
- Trained Logistic Regression, Random Forest, and Gradient Boost models on all seasons in the data set.
- Features included:
  - Overall team winning percentage
  - Winning streak
  - Goals per game
  - Power play goals per game
  - Shots per game
  - Hits per game
  - Penalty minutes per game
  - Giveaways/takeaways per game
  - Time travel


Cross-Validation Scores
Logistic Reg: 0.5486
Random Forest: 0.5471
Gradient Boost: 0.5520
Scores
0.59
0.58
0.57
0.56
0.55
0.54
0.53
s_20122013
s_20132014
s_20142015
s_20152016
s_20162017
s_20172018
s_20182019
Seasons

Models – One Season

- Trained Random Forest and Gradient Boost on 1 season.
- No features really jump out
- Next, trained the models on one season and predicted the others.
- Dropped time travel




Cross-Validation Scores
Random Forest: 0.5549
Gradient Boost: 0.5439
0.60
0.58
0.56
0.54
0.52
Scores
s_20122013
s_20132014
s_20142015
s_20152016
s_20162017
s_20172018
s_20182019
Seasons

Conclusions

- Don’t bet money based on this model.
  - Simulated 100 bets of $100 on the favored team using this model with 55% accuracy and lost about $1000 on average.
  - Need about 60% accuracy to just break even.
- Hockey is difficult to predict.
  - Winning percentage ranged 40% - 60%, so decent amount of parity over all seasons.
- Games closer than 2 goal spread suggests.
  - Often the last goal is an empty net goal, so goal spread probably closer to 1.0 realistically.
- Puck luck

Further Research

- Add in player statistics as features, starting with the goalie
  - Lineups and injuries will matter
- Use a neural network
- Regression – predict goals per game or goal spreads

Sources

- 1. https://www.legalsportsbetting.com/how-much-money-do-americans-bet-on-sports/
- 2. https://www.corsicahockey.com/corsica-predictions-explained