

NHL Game Prediction Modeling

By Gary Schwaeber



Overview



With sports betting becoming increasingly popular and mainstream, data science can be used to make superior decisions over gut intuitions.

- Moneyline is king in NHL betting
- How do you gain an edge?
 - You know the truer probability of game outcomes vs implied odds from the moneyline
- Goal: Build a model that outputs more accurate probabilities
 - Profitable over long time frame



Data Collection



Game Log Stats

Scraped from
Natural Stat Trick



Official Results

Scraped from NHL
API via
hockey-scraper



Seasons

2016-2017 to Present

Features



5v5 Score and Venue Adjusted

- Fenwick For %
- xGF %
- Goals For %
- Shooting %

Elo Rating

- Measure a team's strength over time

Back to Back

- Was the home or away team playing back to back?

Powerplay

- xGF per minute PP
- Goals for per minute PP

Penalty Kill

- xGA per minute PK
- Goals Against per minute

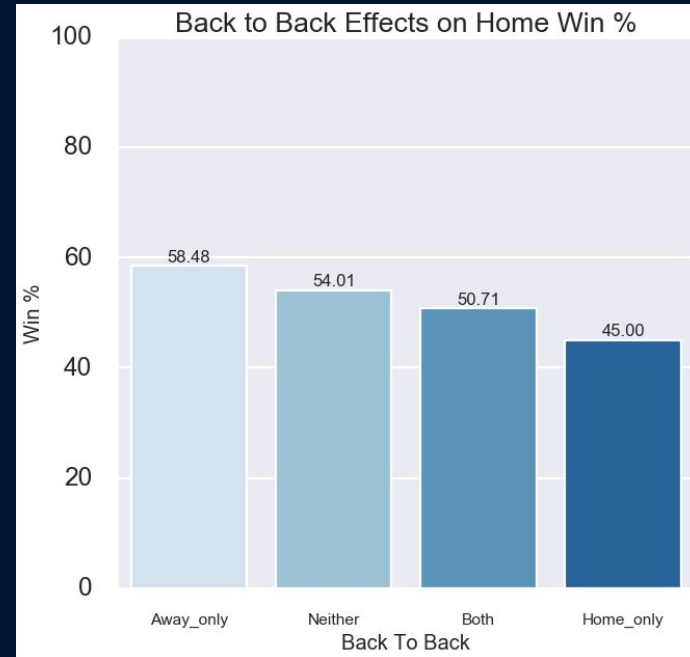
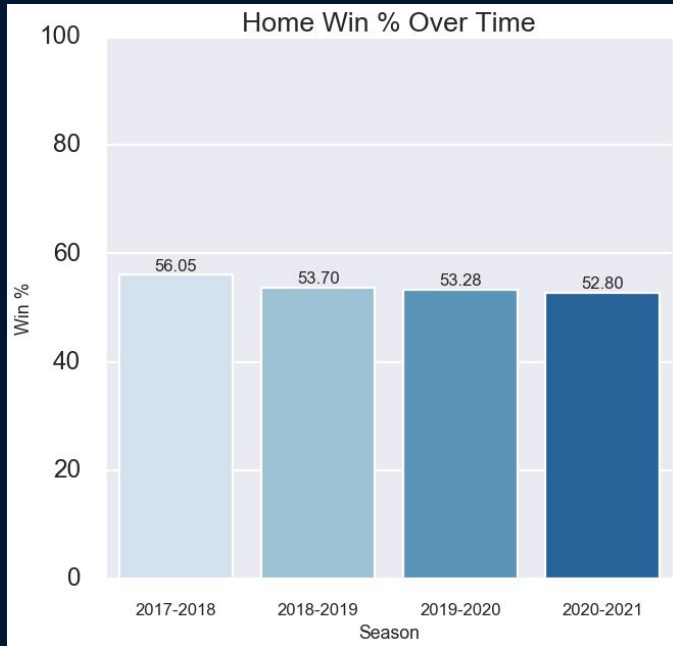
Goaltending

- Fenwick Save %
- Goals Saved Above Expected Per 60
- High Danger Save %

Home Win % and Back to Back Effects

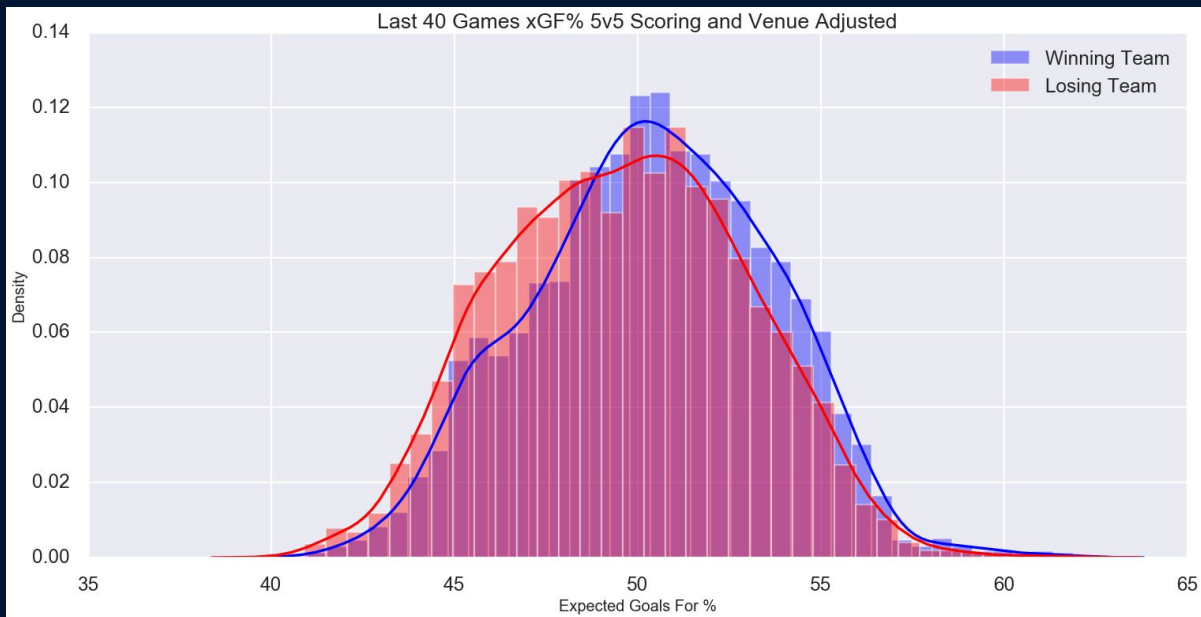


- Home ice advantage, but declining
- Playing B2B affects home team more



Goals For %: Winning vs Losing Team

- A lot of overlap
 - Good teams lose and bad teams win often
- Any feature indicating team strength will only slightly nudge win probability



Modeling Process



Optimize For Log Loss

- Lower is better
- View Accuracy



Train/Test Split

- Train on seasons 17-18 to 19-20
- Test on 20-21



Models

- Logistic Regression
- AdaBoost
- Gradient Boosting
- Neural Network



Features

- 5 and 40 Game Rolling
- 40 Game Rolling
- All Rolling windows (3,5,10,20,30,40) with RFECV



Iterate

- Use Grid Search to tune hyperparameters



Results

- 40 Game Neural Network scored best
- 40 Game Rolling only feature set generally performed best
- Best model is competitive with other publicly published models

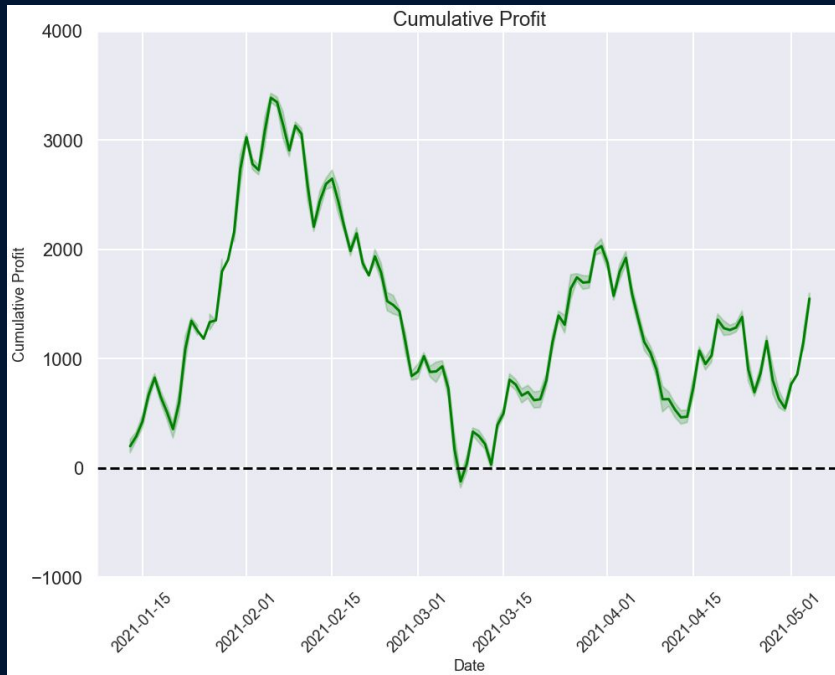
	Training CV Accuracy	Training CV Log Loss	Test Accuracy	Test Log Loss
40 Neural Network	0.5773	0.6726	0.6024	0.6555 
40 Logistic Regression	0.5787	0.6742	0.6024	0.6568
5 and 40 Neural Network	0.5843	0.6736	0.6037	0.6579
40 AdaBoost	0.5723	0.6758	0.6146	0.6605

2021 Season Competitor Results

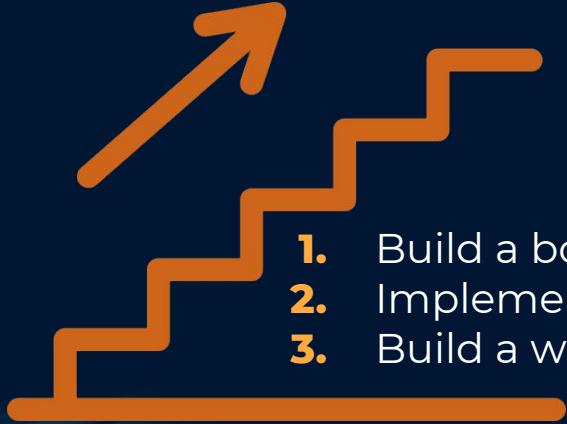
Model	Log Loss
The Athletic	0.6493
Implied odds	0.6517
Hockey-Statistics	0.6531
BayesBet	0.6556
BulsinkBot	0.6599
MoneyPuck	0.6620
TopDownHockey	0.6678
HockeyViz	0.6728
MoreHockeyStats	0.6781

Was The Model Profitable?

- Back tested against historical odds
- Strategy: bet to win \$100 on the team where the model's probability to win is greater than the implied odds
- 2.04% ROI per bet
- 59.1% Home / 23.2% Away / 17.6% No Bet
- 47.2% Underdog / 23.2% Favorite / 17.6% No Bet



Next Steps



- 1.** Build a bottom up model using player based data
- 2.** Implement a voting classifier
- 3.** Build a web app to publish future predictions

THANKS!

Do you have any questions?
Gary Schwaeber
gschwaeb@gmail.com



linkedin.com/in/gary-schwaeber



github.com/gschwaeb



gsstats.medium.com

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

