

Boston Marathon

Cheaters and Winners

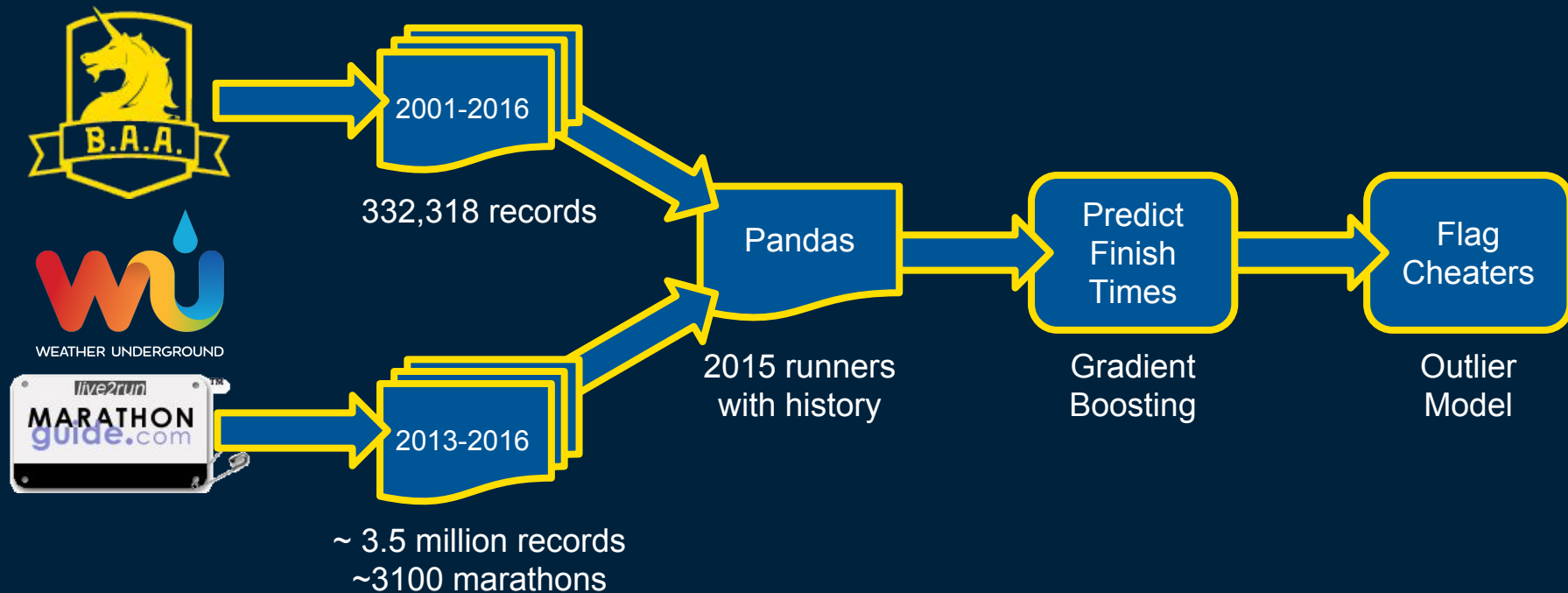
Timothy Chan
trchan@gmail.com
github.com/trchan/boston-marathon

Not "Everyone's a Winner"

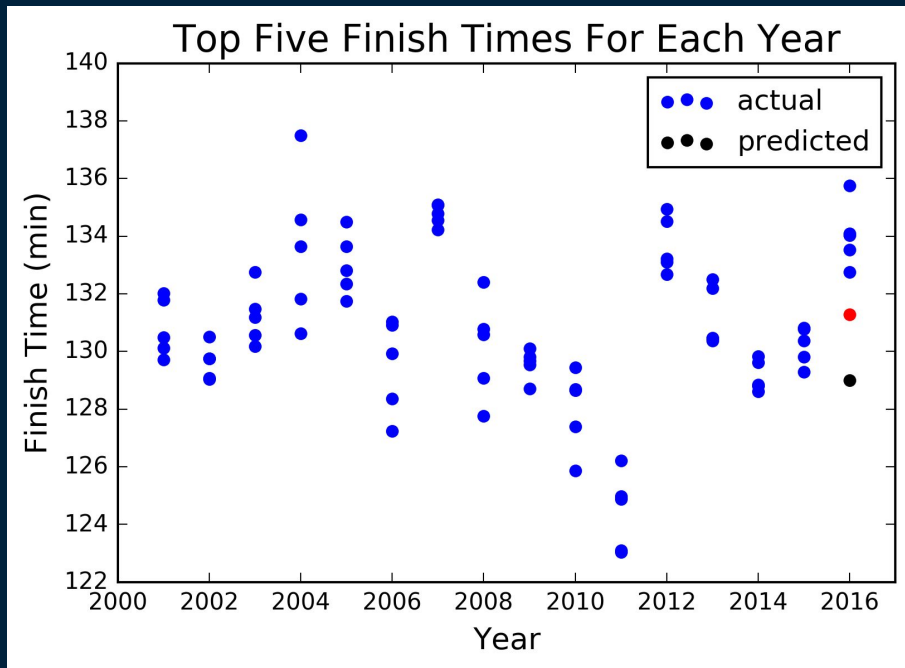
Mike Rossi



The Data Pipeline



Predicting The Winning Time



Timothy Chan
@trchan1

Boston Marathon
@bostonmarathon

#Dat Lemi Berhanu Hayle (🇪🇹ETH) is your 2016 #BostonMarathon men's champion with an unofficial time of 2:12:44 🏁🏆

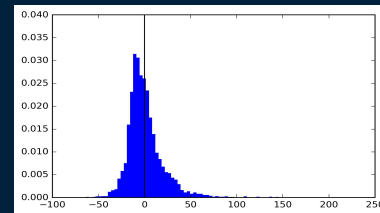
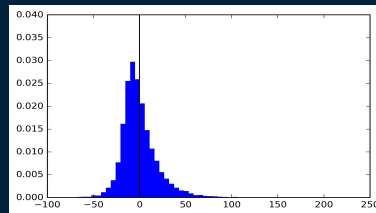
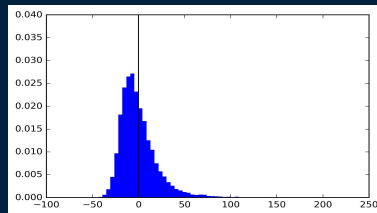
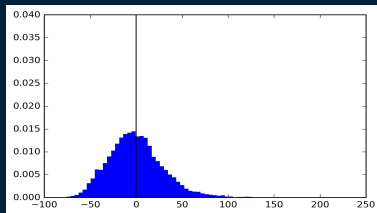
RETWEET 1

11:19 PM

RETWEETS 434 LIKES 359

9:14 AM - 18 Apr 2016

Predicting Finish Times



Model

Mean

Linear Regression
w/bib number

Linear Regression
w/running history

Gradient Boosting
w/running history

R^2

--

0.617

0.660

0.809

**Standard
Error**

37:42 min

25:24 min

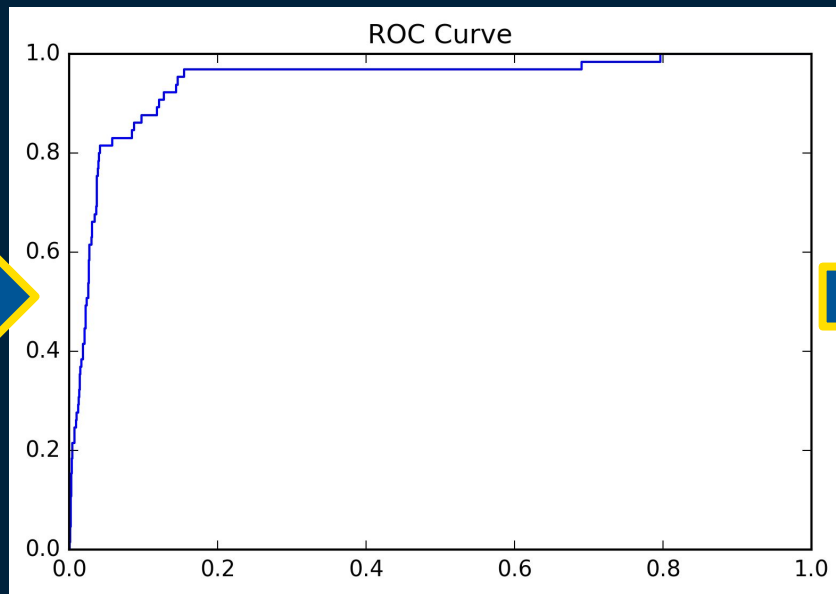
20:21 min

16:29 min

Outliers = Cheaters

**Marathon
Investigation.com**

58 of 28,000
(0.2%)



Score = 0.938



**32 min
threshold**

53 of 1261
(4.2%)

Difficulties Encountered

- Matching running records across different databases
- No obvious source for cheater labels
- Engineering an injured feature

Future Work

- Find cheaters in the qualifying marathons
- Get list of 2016 outliers inspected

Thank You

Timothy Chan

trchan@gmail.com

github.com/trchan/



