

The Making of a Real-World Moneyball

Finding Undervalued Players with H₂O, LIME and Shiny



Jo-fai (Joe) Chow

Data Science Evangelist /
Community Manager

joe@h2o.ai

@matlabulous

More Info → [https://bit.ly/](https://bit.ly/satRdayAMS_moneyball)

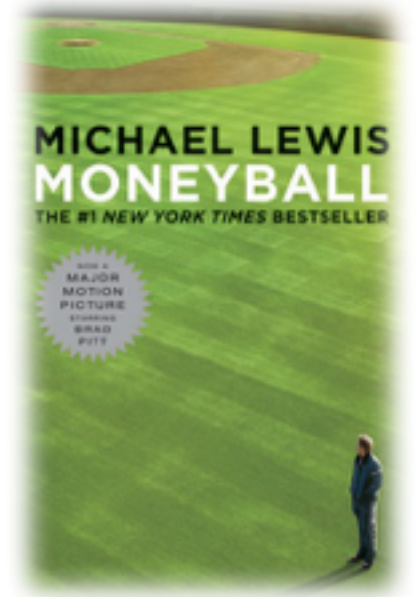
satRdayAMS_moneyball

Moneyball: The Multimillion-Dollar Business Problem

The quest to find the most undervalued players
(before other teams notice them)



Source: Moneyball, 2011 Columbia Pictures



The Real Business Problem in Major League Baseball (MLB)

- Existing Forecasts (e.g. ESPN) are usually projections for the next year only.
- MLB players usually consider terms for 3 to 5 years when they sign a new contract.
- MLB teams need to consider players' long-term performance (i.e. > 1 year).



The Moneyball Team



David Kearns

PM @ IBM Data Science



Ari Kaplan

Mr. Moneyball @ Aginity



Jo-Fai Chow

Data Scientist @ H₂O.ai

In case you're wondering... final project result

\$20M

trade 2 weeks prior to the
season beginning



Baseball Player Performance Data

- Open data – Lahman Database.
- Proprietary data from Ari Kaplan – our real Moneyball guy.
- Enriched Lahman data with Ari's Data



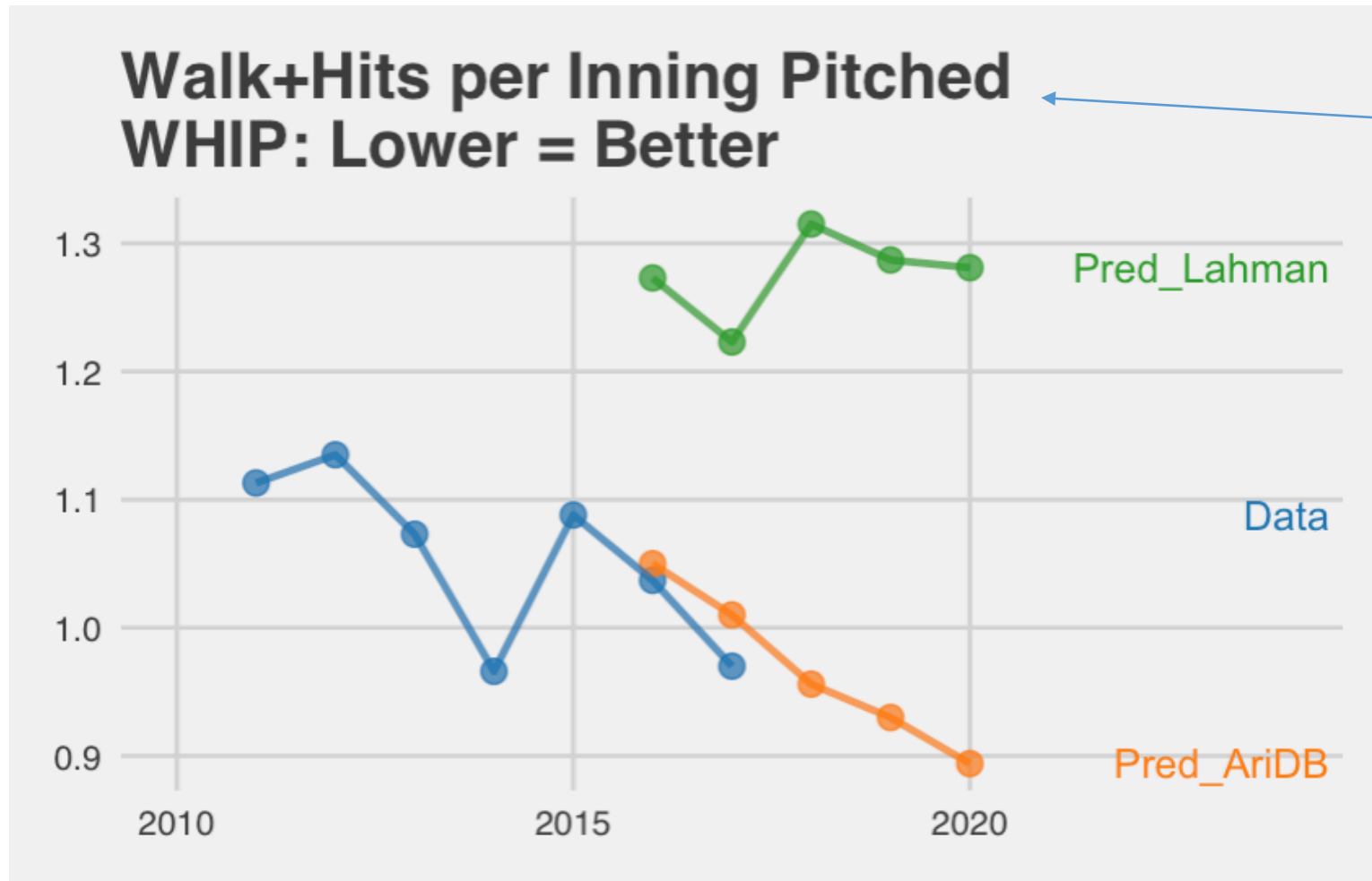
Predictive Modelling – H₂O AutoML

- Framed data as regression problems for performance prediction.
- Historical player performance as features.
- Used H₂O AutoML to build ensembles (linear model, random forests, gradient boosting, and deep neural networks).



```
# Install 'h2o' from CRAN  
install.packages('h2o')
```

Predictive Modelling – H₂O AutoML

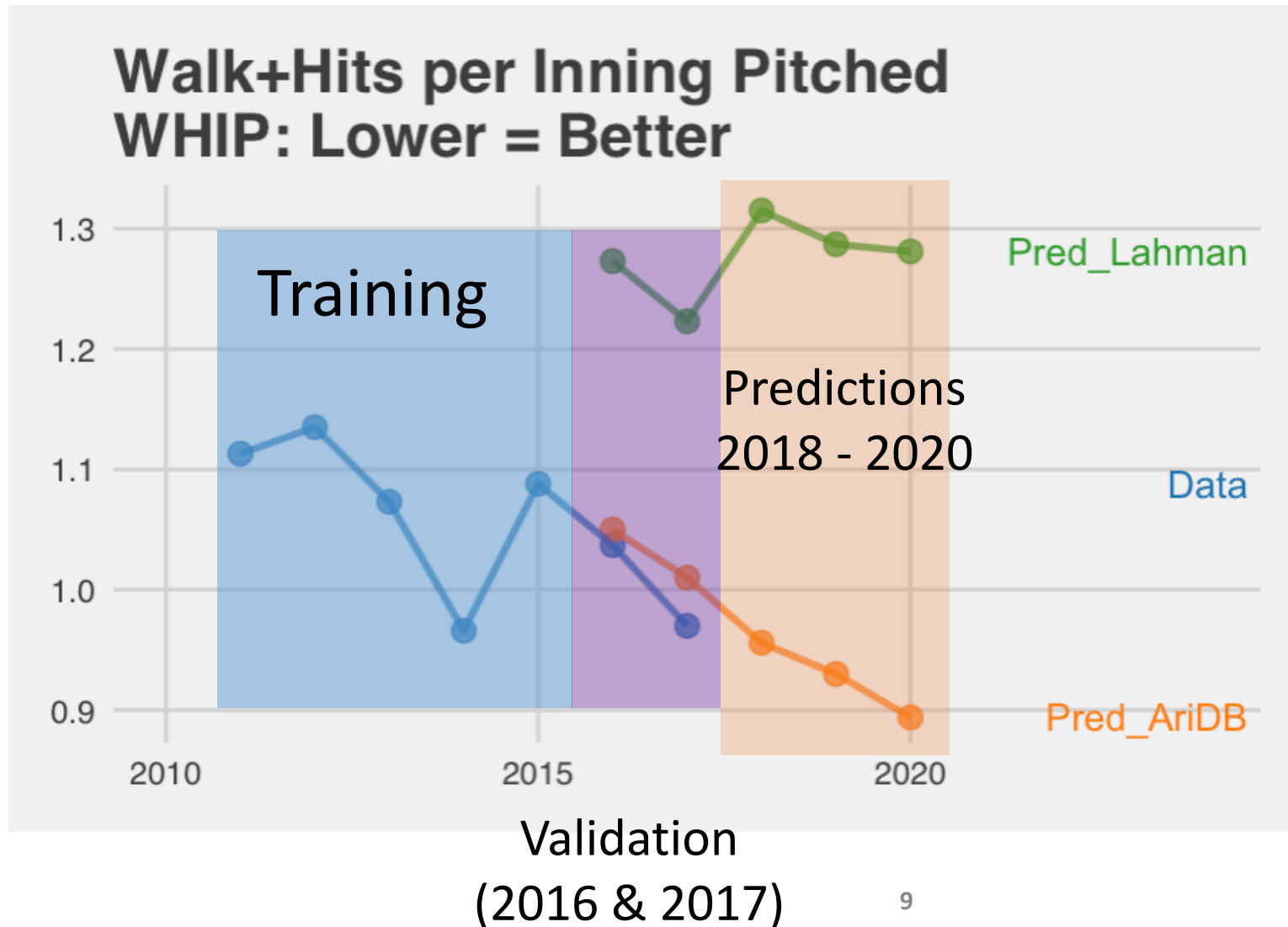


One of Many Targets
(e.g. Home Runs, Batting Average)



```
# Install 'h2o' from CRAN  
install.packages('h2o')
```

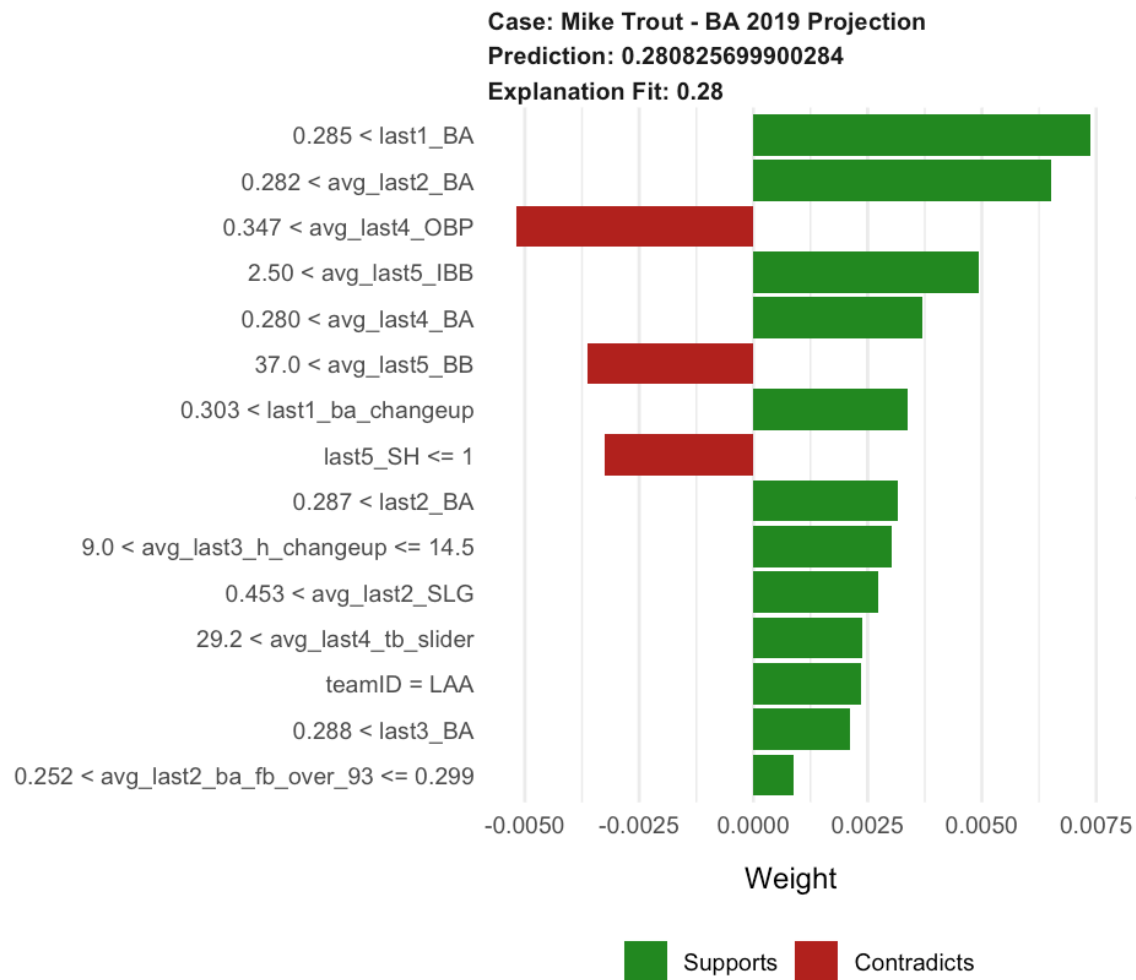

Predictive Modelling – H₂O AutoML



```
# Install 'h2o' from CRAN  
install.packages('h2o')
```

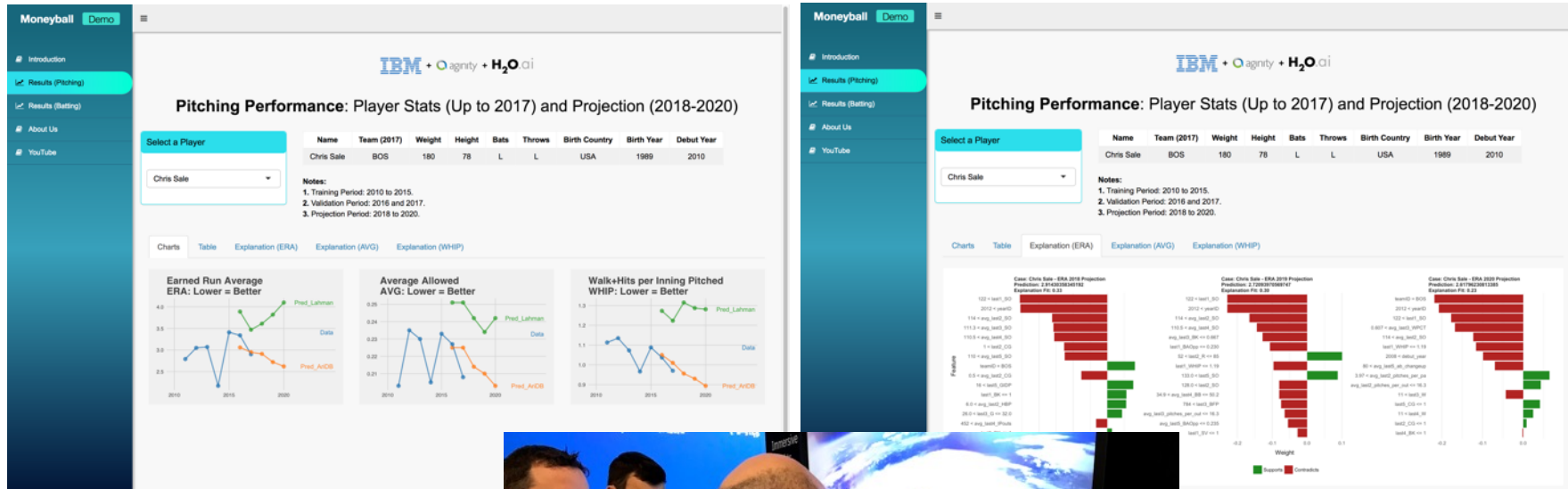
Explaining the Predictions

LIME – Local Interpretable Model-agnostic Explanations



```
# Install 'lime' from CRAN  
install.packages('lime')
```

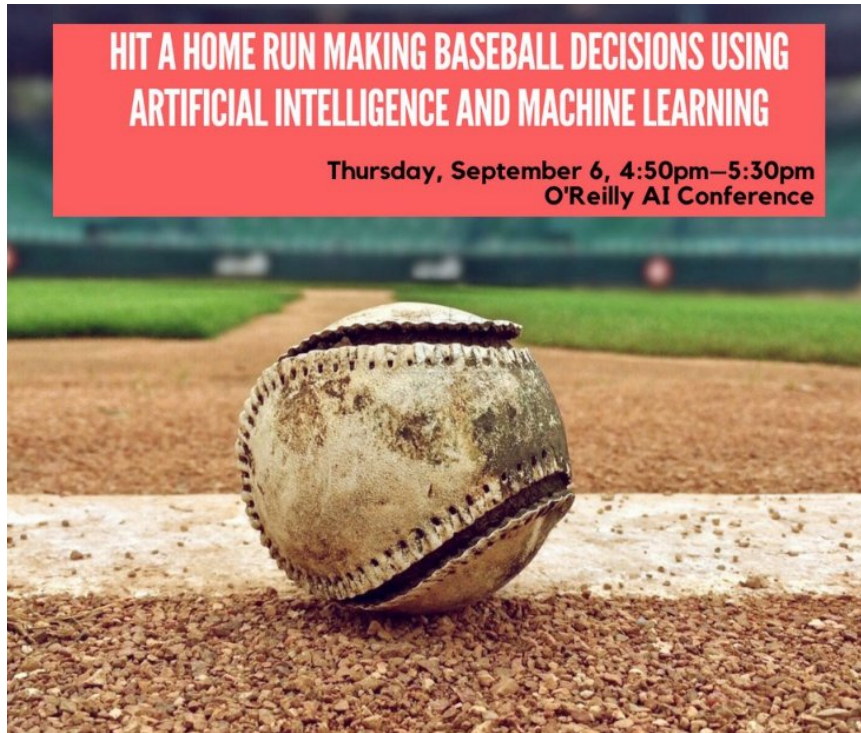
Putting Everything Together – Moneyball Shiny App



Live Demo



Full Presentation
if you want to hear from Ari himself



6th Sep, San Francisco



30th Oct, London

Thanks!



- More Info, Code, and Slides
 - bit.ly/satRdayAMS_moneyball
- Contact
 - joe@h2o.ai
 - [@matlabulous](#)
 - github.com/woobe