Is a Fastball or Non-fastball Coming?

Chien Yuan Chang

10/29/2021

0.434 s

A 95 mph ball passes 60.5 feet

The Science of Swing

400 ms Reaction Time

From the pitcher releases the ball to the ball reaches the homeplate

Pitch → **Contact or Not**

100 ms Look

From the eyes to the brain

75 ms Think

The brain processes the information and gauge the speed, type and location of the pitch

50 ms Decide

- 25 ms to decide whether to swing
- 25 ms to pick a swing pattern

25 ms Act

The brain sends signal to the body

150 ms Swing

- Complete the swing
- The batter can stop during the first 50 ms

Solution

A classification model to predict which pitch type a pitcher will pitch under certain game situations

Information source: The science of swing, Robert Adair, Yale physicist

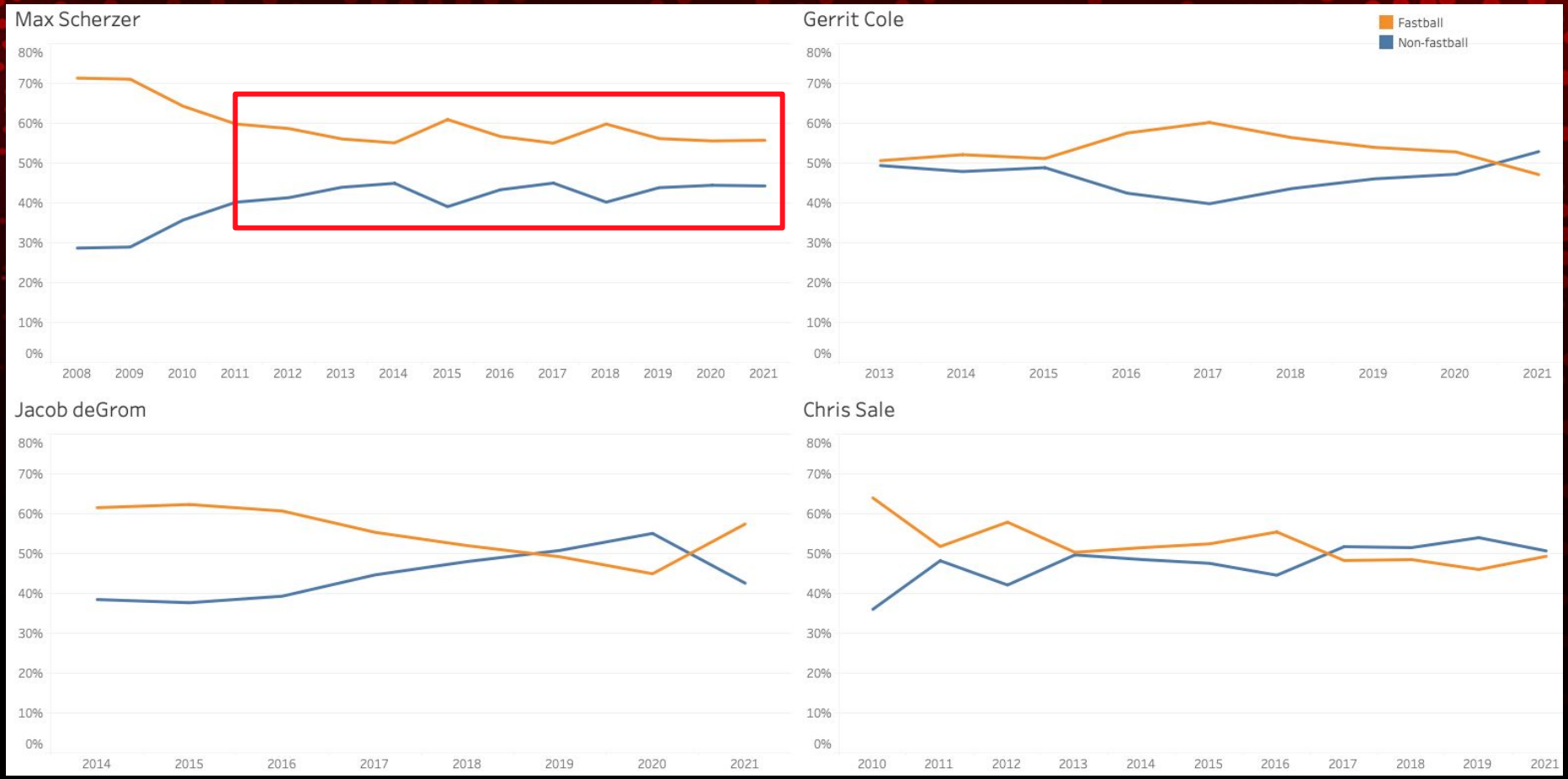
The Target Pitcher

Max Scherzer

- 14-year MLB pitcher
- 8-time MLB All-Star
- 2-time Cy Young Awards
- 2 no-hitter games
- 1 World Series championship
- 3,000 strikeouts

Is a Fastball or Non-fastball Coming?

More stable fastball usage



Data

2015-2019 MLB All Pitch-by-pitch Data

- Downloaded from Kaggle
- Total 3,595,944 pitches
- 66 columns when merging with game and bat data

Data Cleaning and Feature Engineering

- missing and incorrect data in2019
- 13,454 pitches of Max Scherzer (2015-2018)
- Created over 50 new features

Final Model (Metric: Accuracy)

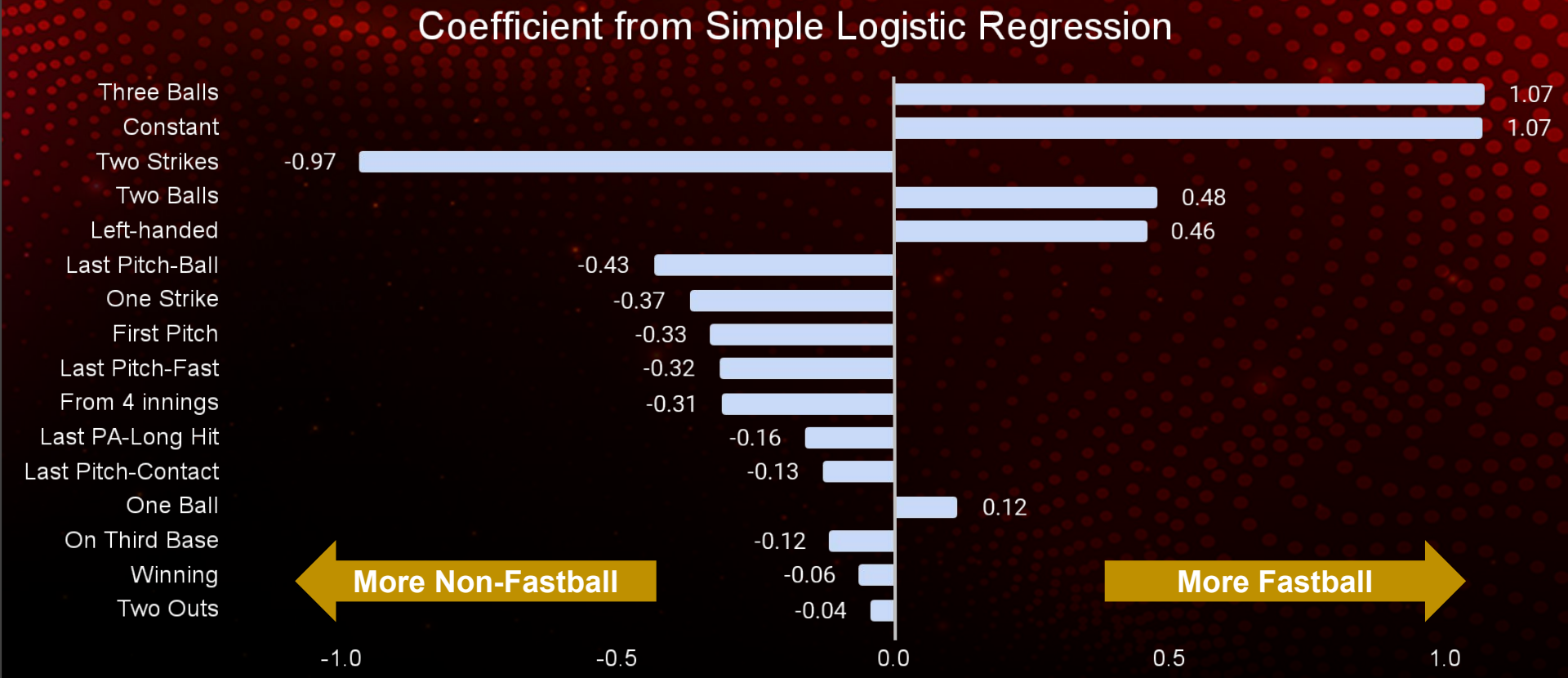
2008-2021 By Pitcher Pitch-by-pitch Data

- Web-scraping from Baseball Savant
- 41,277 pitches of the career of Max Scherzer
- 14 columns
- Supplemental data for EDA

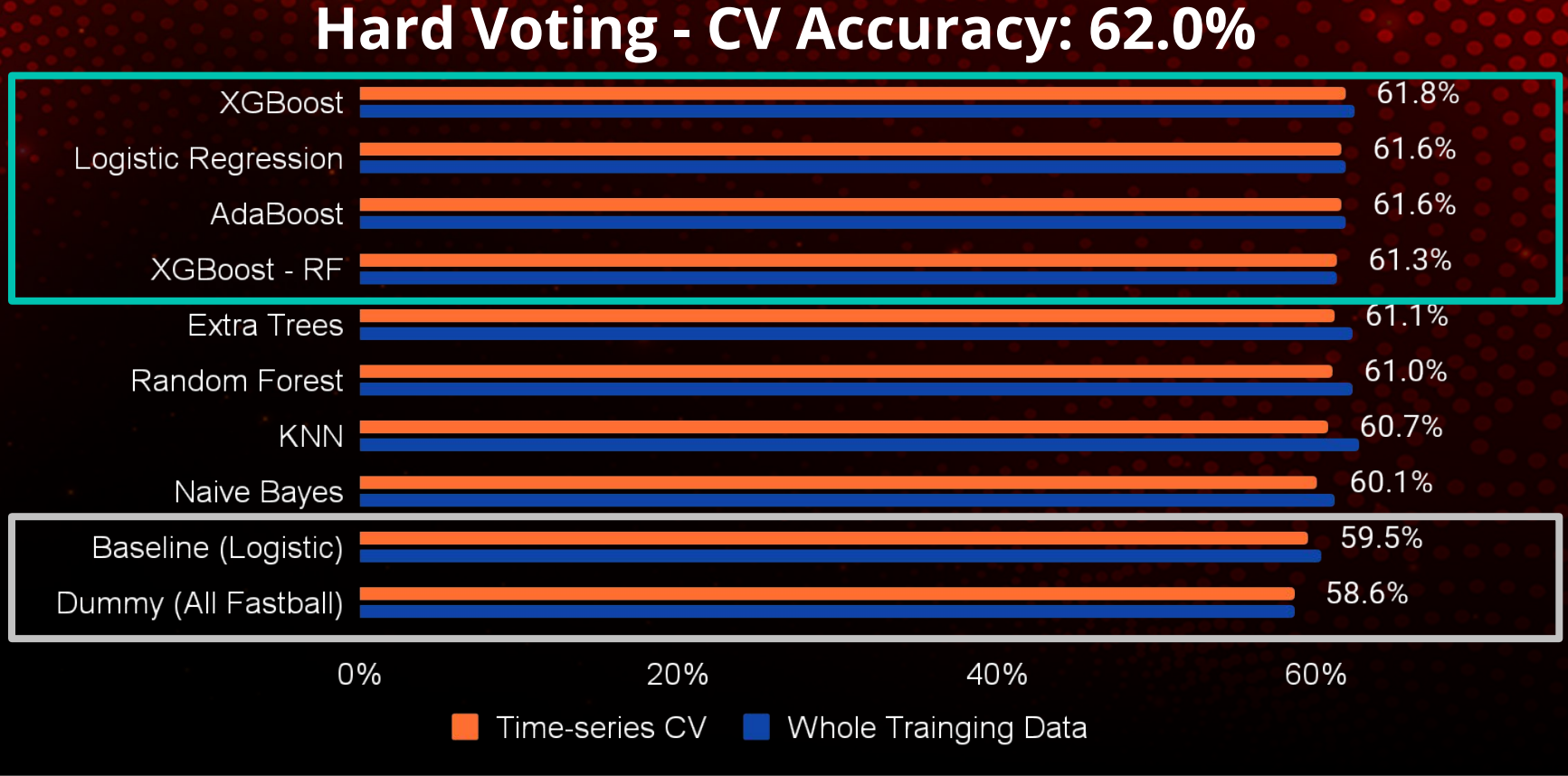
Model Training

- Time-series Split/CV
- Logistic Regression
- KNN
- Random Forest
- XGBoost/AdaBoost
- Naive Bayes
- Voting/Stacking

Final Features



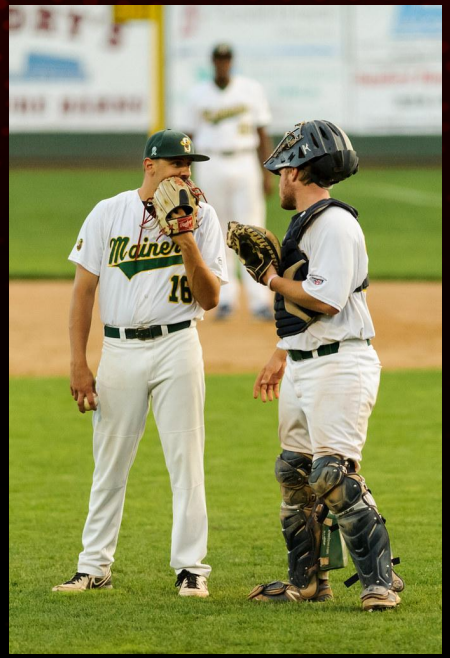
Accuracy by Algorithms on Training Data



Final Model on Test Data
Hard Voting with 4 Algorithms

Accuracy 60.7%		Predicted		
		Fastball (76.1%)	Non-Fastball (23.9%)	**Recall**
Actual	Fastball (58.0%)	47.4%	10.6%	**81.7%**
	Non-Fastball (42.0%)	28.7%	13.3%	**31.7%**
	Precision	**62.3%**	**55.6%**	

A Brain Game



Future Work

- The data of the batters
- The data of the catchers
- More data of years
- More classification on different pitchers
- The precision of a certain pitch type which is easier to become a hit

Thank You!

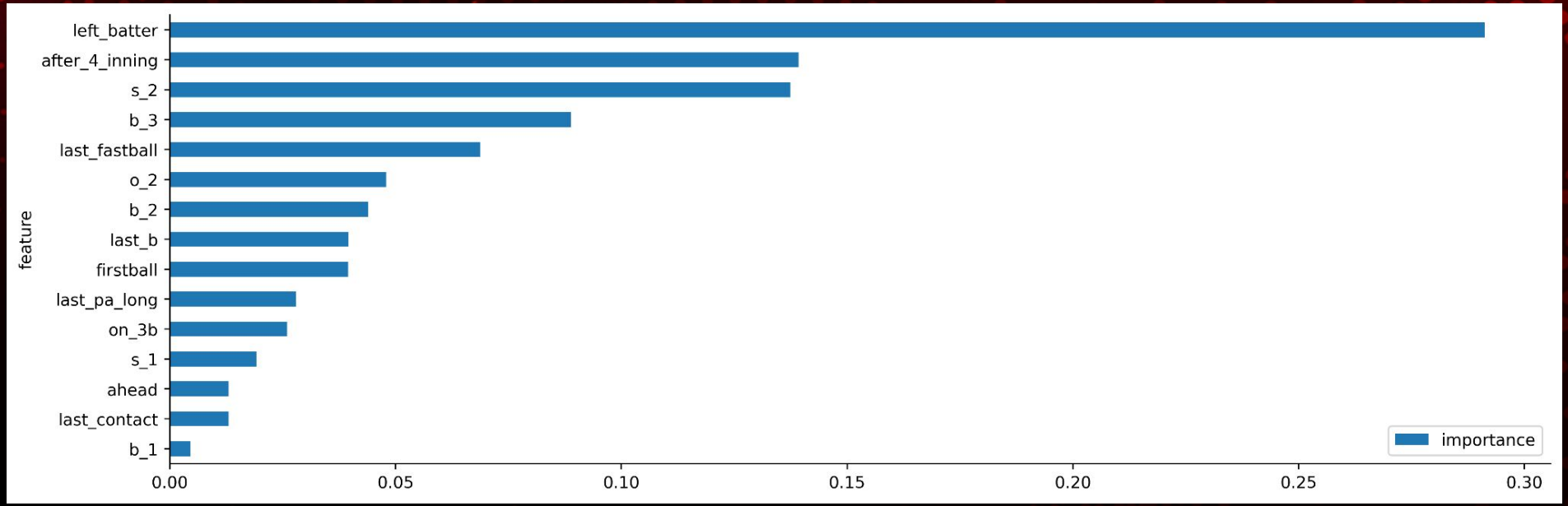
Do you have any questions?

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

Resource

- MLB Pitch Data 2015-2018 on Kaggle
  - https://www.kaggle.com/pschale/mlb-pitch-data-20152018
- Baseball Savant
  - https://baseballsavant.mlb.com/statcast_search
- The Physics of Baseball
  - https://www.amazon.com/Physics-Baseball-3rd-Robert-Adair/dp/0060084367

XGBoost Feature Importance



The Science of Swing

- 400 ms reaction time
- 100 ms from the eyes to the brain
- 75 ms for the brain to process the information and gauge the speed, type and location of the pitch
- 25 ms to decide whether to swing
- 25 ms to pick a swing pattern
- 150 ms to complete the swing

Information source: The science of swing, Robert Adair, Yale physicist