MLB Pitch Classification

• • •

By: Zach Potthoff and Alex Rimerman

What Pitches Do

- Goals of pitching:
  - Throw strikes
  - Make the hitter chase
  - Miss barrels
  - Get quick outs

Gif of Jacob Degrom slider

Gif of Joe Smith slider

- How pitchers do this:
  - Throw pitches that look the same out of the hand (Tunnel)
  - Make them move in different directions
  - Stand out from other pitchers

MLB Pitch Tracking Technology: Hawkeye

- Hawkeye takes images of pitcher at ball release
- Tracks ball flight from release to home plate
- Measures velocity, spin rate, movement, release height, among other advanced metrics
- All MLB in-game data is then uploaded to Baseball Savant

Image of Hawkeye System MLB Stadium

Overview (What We Did)

- Web scraped MLB pitching data

- Built a Random Forest that can classify pitch types

- Adjusted model and features to solve the problem more accurately

- Attempted extension on grouping similar pitches as a pitcher comparison tool

Inspiration

- At MLB games you see instant pitch identification
- We wanted to see if we could replicate this process accurately and efficiently
- MLB uses a neural network, we wanted to try to simplify the model

Pitch Type at MLB Game

Getting Our Data

- Pybaseball
  - Open source web scraping tool found on GitHub
  - Used this to get data from Baseball Savant
    - Baseball Savant has access to every pitch thrown in the MLB in recent years
- Took every pitch from 2023 season as our data set
- Filtered to only features we anticipated as being important to pitch classification
  - Pitch velocity, horizontal/vertical movement, spin rate

Initial Testing

- Started with 2023 World Series data
  - Began with drastically different pitches (e.g. fastball/curveball) and progressed to more complicated pitch types
  - Saw a high accuracy and efficient model, so decided random forest was a good model to move forward with
- Repeated process for entire 2023 season

Results

- Confusion matrix models 3 easily distinguishable pitches from full 2023 season
- Predicted these pitch types very well with 99.7% accuracy on test set

Fastball/Changeup/Curveball Confusion Matrix for a Right Handed Pitcher
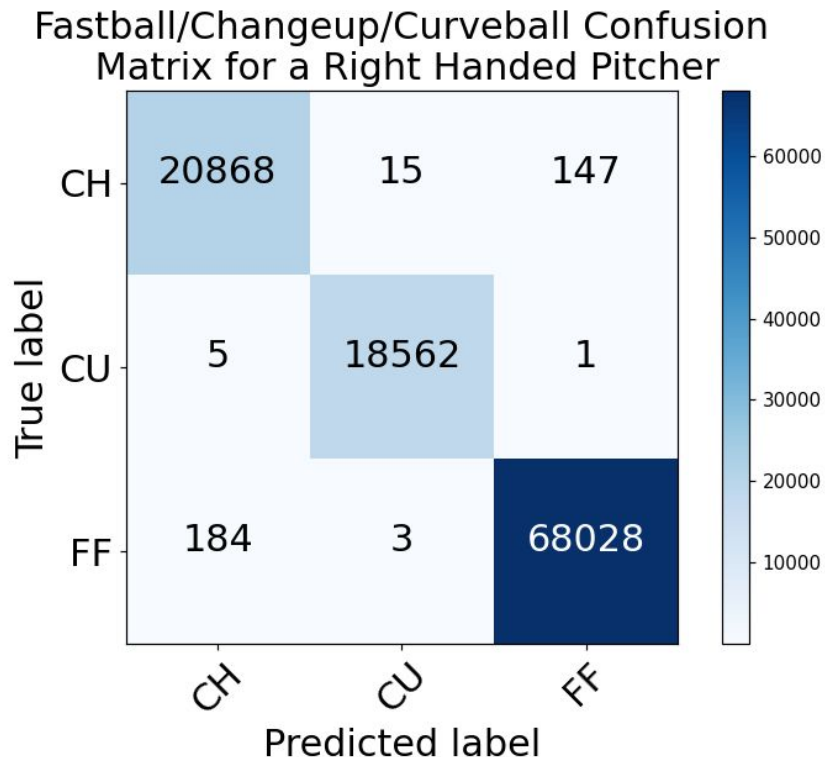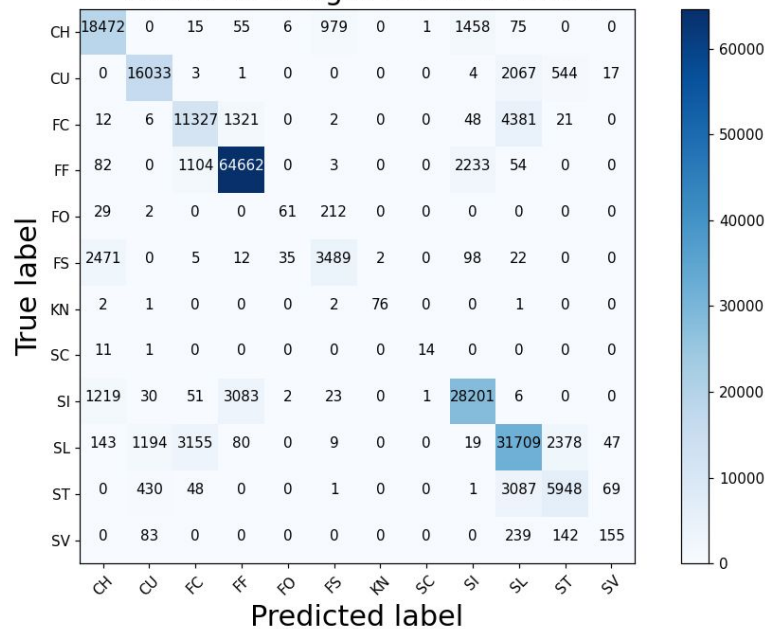
Expansion to full data (84.5% accuracy on both)

All Pitches Without Release Confusion Matrix for a Left Handed Pitcher
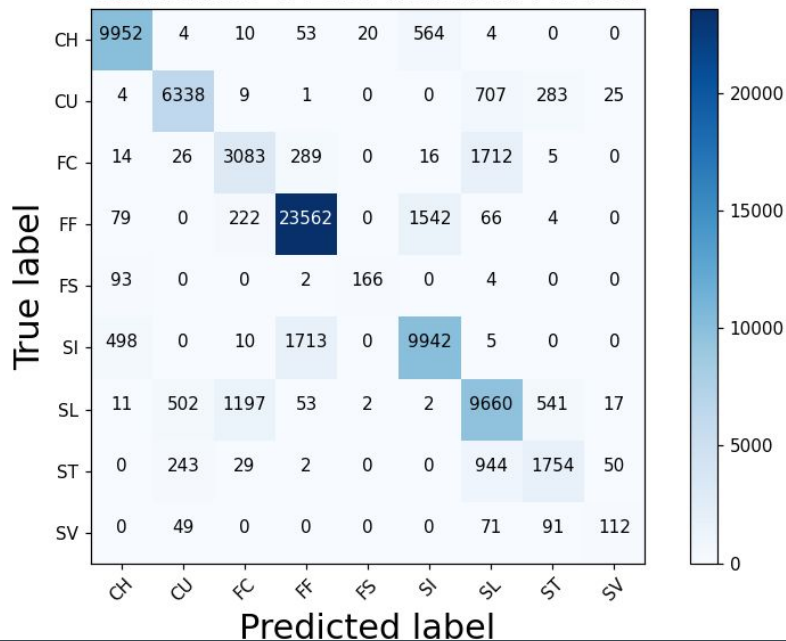
All Pitches Without Release Confusion Matrix for a Right Handed Pitcher

More Features?

- Model accuracy quickly goes down with more pitch options
  - Either missing features or have a bad model
- Looked into potential feature expansion
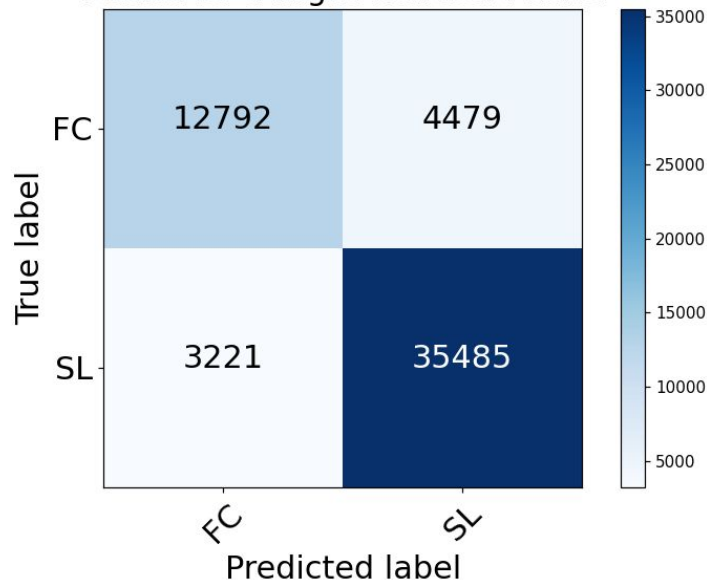 - Considered release point information and reattempted testing
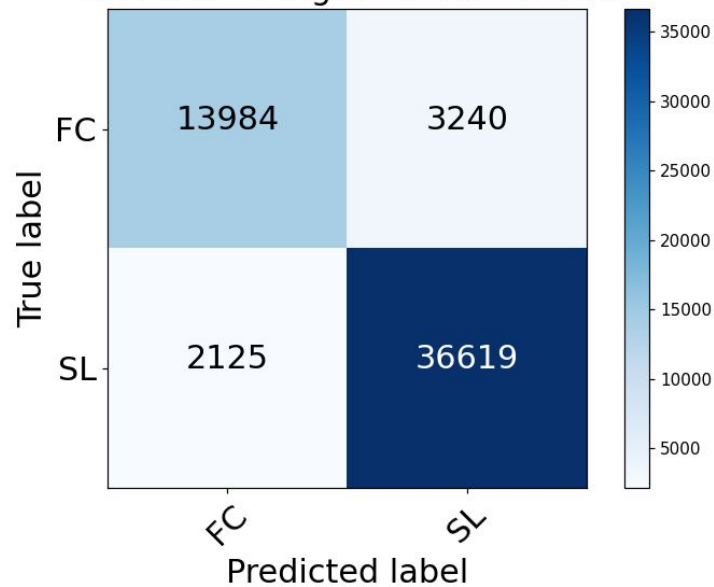

Image of Miles Mikolas Curveball

Image of Jack Flaherty Slider

Results on Slider/Cutter (86.2% vs 90.4% accuracy)

Final Output (93.2% and 90.4% accuracy)

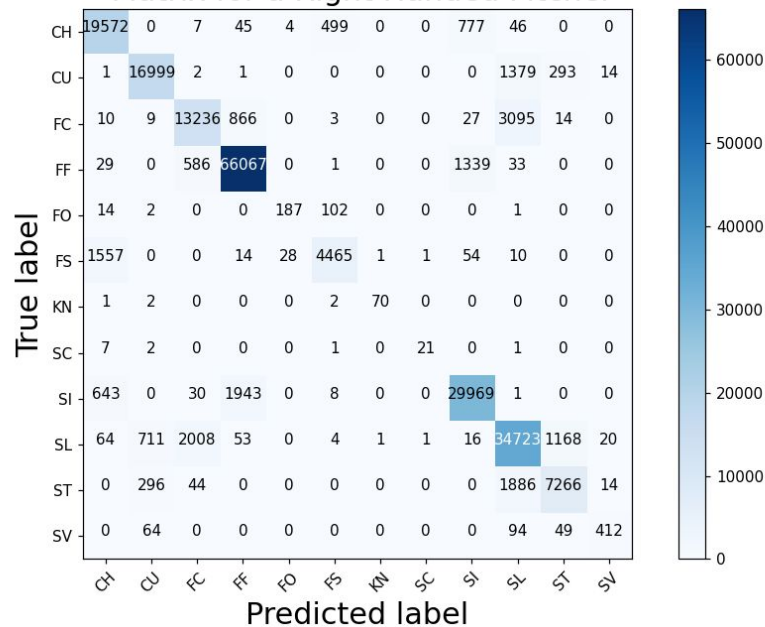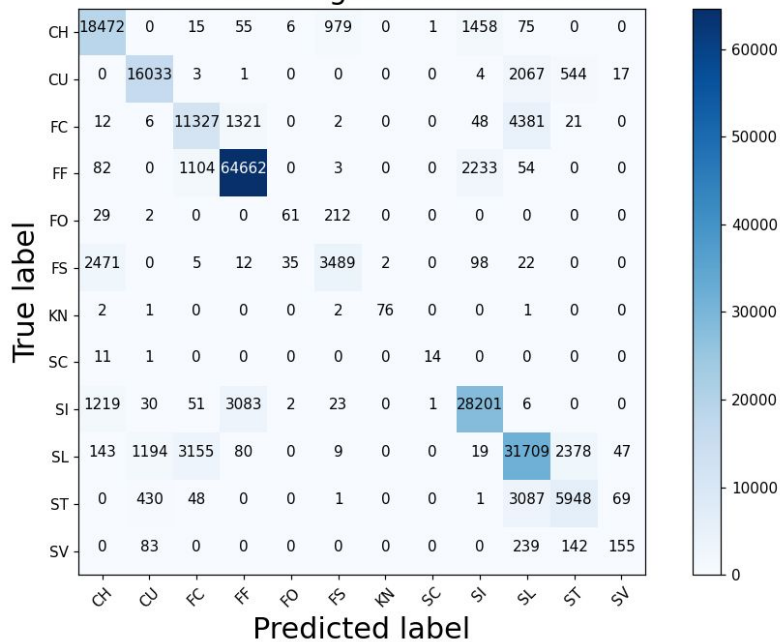All Pitches With Release Confusion Matrix for a Left Handed Pitcher
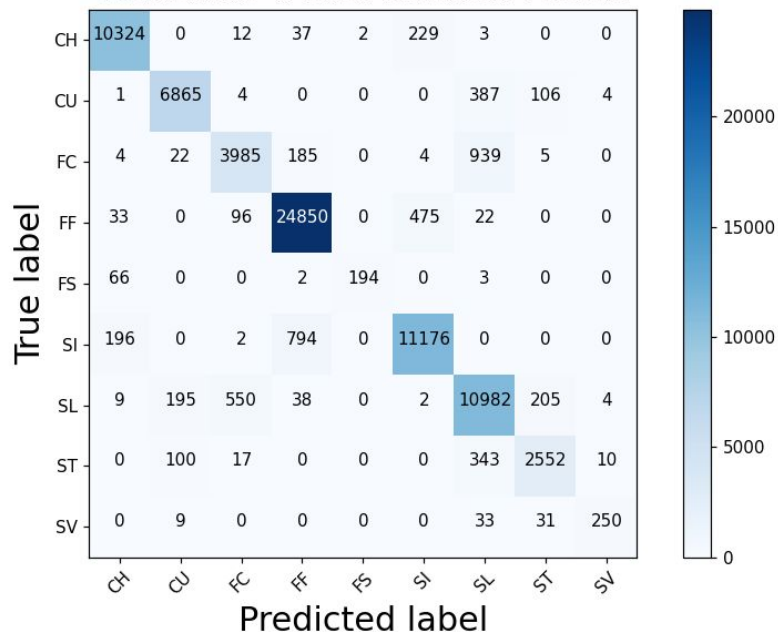
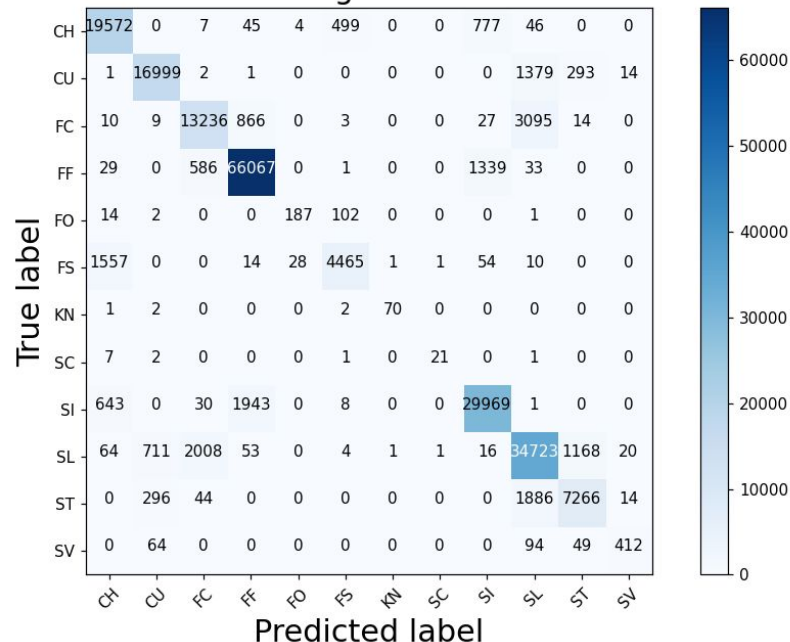All Pitches With Release Confusion Matrix for a Right Handed Pitcher

Compare Old/New (84.5% vs. 90.4% accuracy)

Extension

- Used KNN to find 5 most similar pitches to a given pitch from the data set
- KNN not complex enough to do this
- Can be seen by differing pitch types

```
Pitcher, pitch type
['Bickford, Phil' 'FF']
The following are the 5 'nearest' pitches and who threw them.
['Medina, Luis' 'SL'] ['Gray, Jon' 'SL'] ['Barlow, Scott' 'SL']
['Romano, Jordan' 'SL'] ['Williams, Trevor' 'FF']

Pitcher, pitch type
['King, Michael' 'CH']
The following are the 5 'nearest' pitches and who threw them.
['Machado, Andrés' 'CH'] ['Olson, Reese' 'SL'] ['Greene, Hunter' 'SL']
['Gausman, Kevin' 'FF'] ['Bassitt, Chris' 'FC']

Pitcher, pitch type
['Gray, Sonny' 'CU']
The following are the 5 'nearest' pitches and who threw them.
['Eovaldi, Nathan' 'FC'] ['Eflin, Zach' 'SI'] ['Uribe, Abner' 'SL']
['Webb, Jacob' 'FF'] ['Gallen, Zac' 'SL']

Pitcher, pitch type
['Campbell, Isaiah' 'ST']
The following are the 5 'nearest' pitches and who threw them.
['Lugo, Seth' 'CH'] ['Hendriks, Liam' 'FF'] ['Martin, Chris' 'FC']
['Hendricks, Kyle' 'SI'] ['Brash, Matt' 'SL']
```

Wrap-Up

- Achieved goal of 90% accuracy
  - Extrapolated to better than 1 second runtime per pitch
- Advanced model seems required for nearly 100% accuracy, but can significantly reduce complexity for good accuracy
  - Potentially looking into attempting our own neural network
- Found that pitch groupings is not as simple as a linear problem
  - Work to be continued in this area

Thank you!

• • •

Any Questions?

Image Citing

- https://technology.mlblogs.com/introducing-statcast-2020-hawk-eye-and-google-cloud-a5f5c20321b8 (Hawkeye Image)
- Google GIF and Stickers feature
- https://technology.mlblogs.com/mlb-pitch-classification-64a1e32ee079 (Pitch Type Scoreboard)