Risk Analysis: Predicting Tommy John Surgery

John Streaman

Business Understanding

Collect data on MLB pitchers to anticipate risk for TJ surgery before signing, & mitigate practices that may result in needing surgery.

Intro: Who is Tommy John?

First player to receive surgery on UCL (1974)

Replace with tendon from persons forearm/ hamstrings

- (sometimes cadaver)
- Before this, considered career ending injury (dead arm)

- Had a successful career (26 total seasons)

Protect Your Investment

Signed record deal - 10 Year, $700 Million

Potential Fees:

- Luxury tax
- (International signing fees)
- Insurance
- Ticket sales, merch, etc.

Prior TJ Surgery 2018

September 2023

- Shutdown for injury
- MRI - torn UCL
- Underwent 2nd TJ Surgery

Shohei Ohtani injury update: Angels superstar shut down for season, likely getting elbow surgery soon

Ohtani's on-field Angels career could be finished, but he will be around the team for the rest of the season

By Mike Axisa Sep 18, 2023 at 5:13 pm ET • 2 min read

Data

General pitching statistics (1972-2023) from baseball-reference.com

Running Google doc of baseball players who received TJ surgery

- Advanced pitching metrics gathered from pybaseball
- Collected, cleaned, and merged to create a dataset

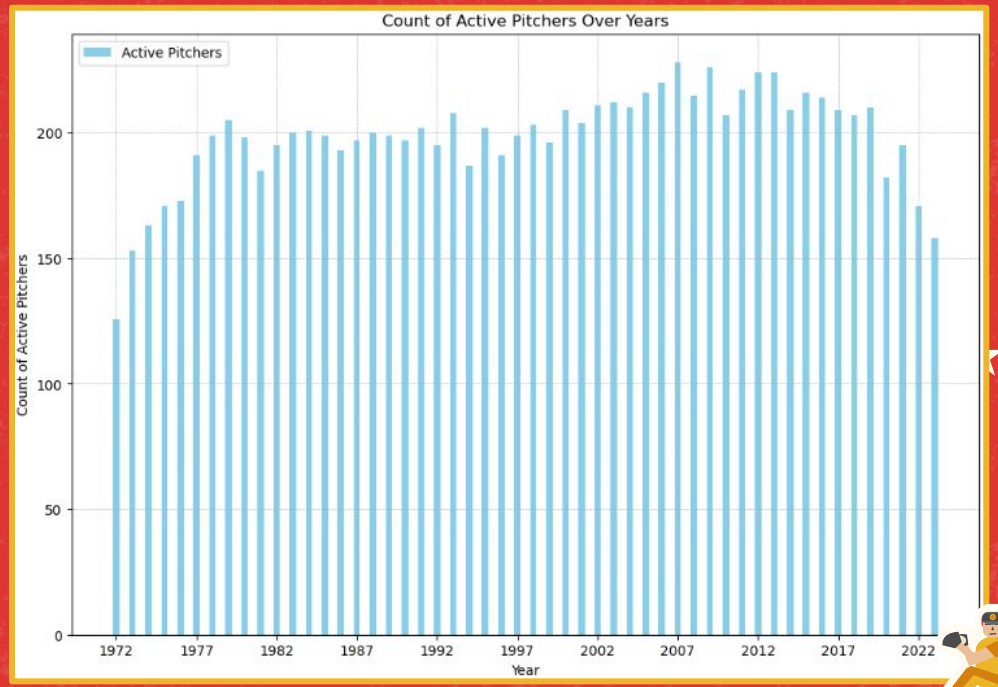
Historic Data

1972 - 2023

Starting Pitchers only

Minimum 320 career innings thrown

- Early 2000's
- Dramatic rise in number of pitchers

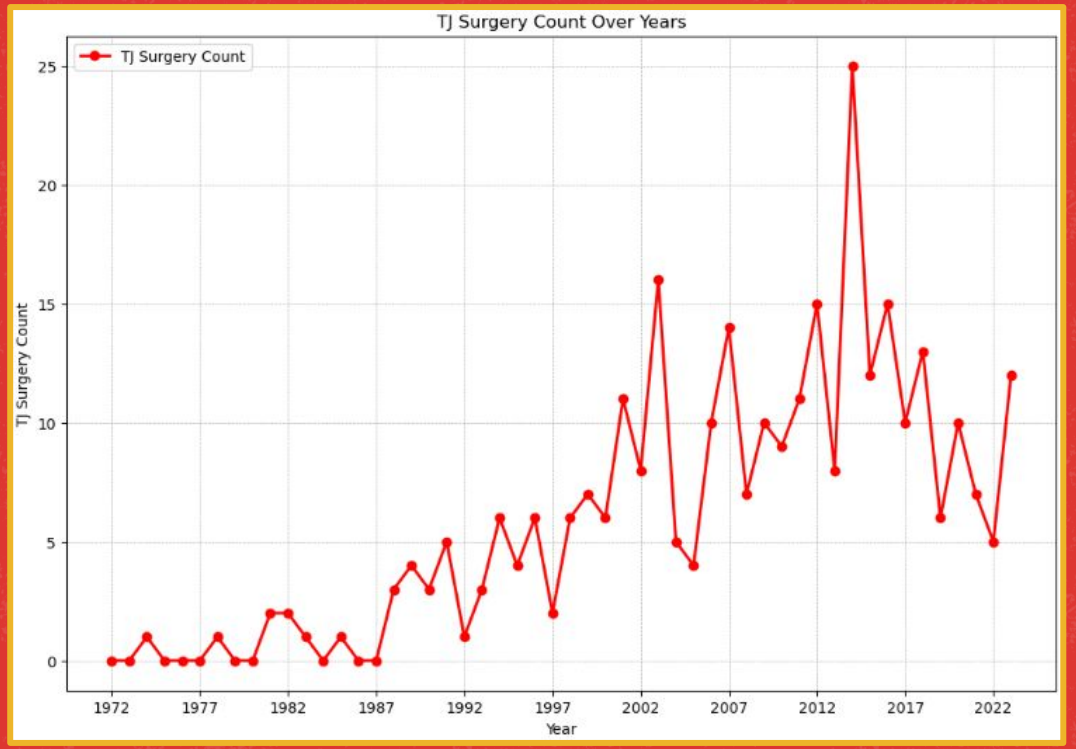


Historic Data cont.

TJ Surgery Count

Dramatic spikes:

- 2003
  - ‘Steroid Era’
- 2014
  - Increase in younger generation play, higher velocity



Advanced Pitching Metrics

Seasonal data - limited

- 2008-2023
- Starting Pitchers, min. 320 innings (career)

Grouped features for each pitcher by season

- Age, season, Throws (L/R)
- Total number thrown for each pitch type
  - fastball, curveball, slider, etc.
  - Avg. speed, velocity, & release position

Features

- 129

Target - ‘Surgery’

- Binary (1.0 - yes, 0.0 - no)
- Count imbalanced
  - 1.0 - 916
  - 0.0 - 2772

Pitch Types

Fastball

- Fastball (FF)
- Sinker (SI)
- Cutter (FC)

Breaking ball

- Curveball (CU)
- Knuckle-curve (KC)
- Slow curve (CS)

Other

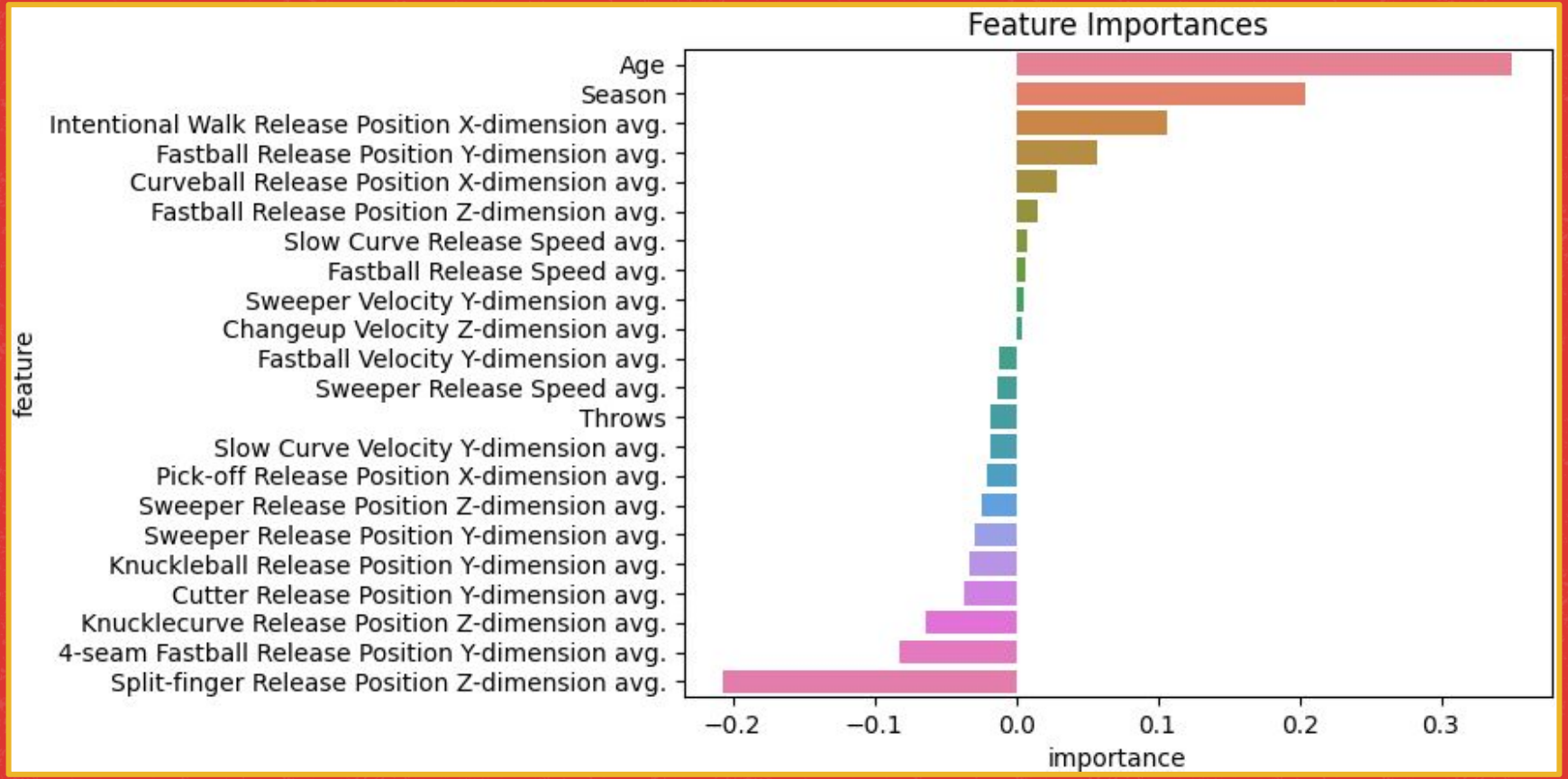
- Eephus (EP)
- Absent (AB)
- Intentional Walk (IN)

Off-speed

- Change-up (CH)
- Split-finger (FS)
- Forkball (FO)
- Screwball (SC)

Slider

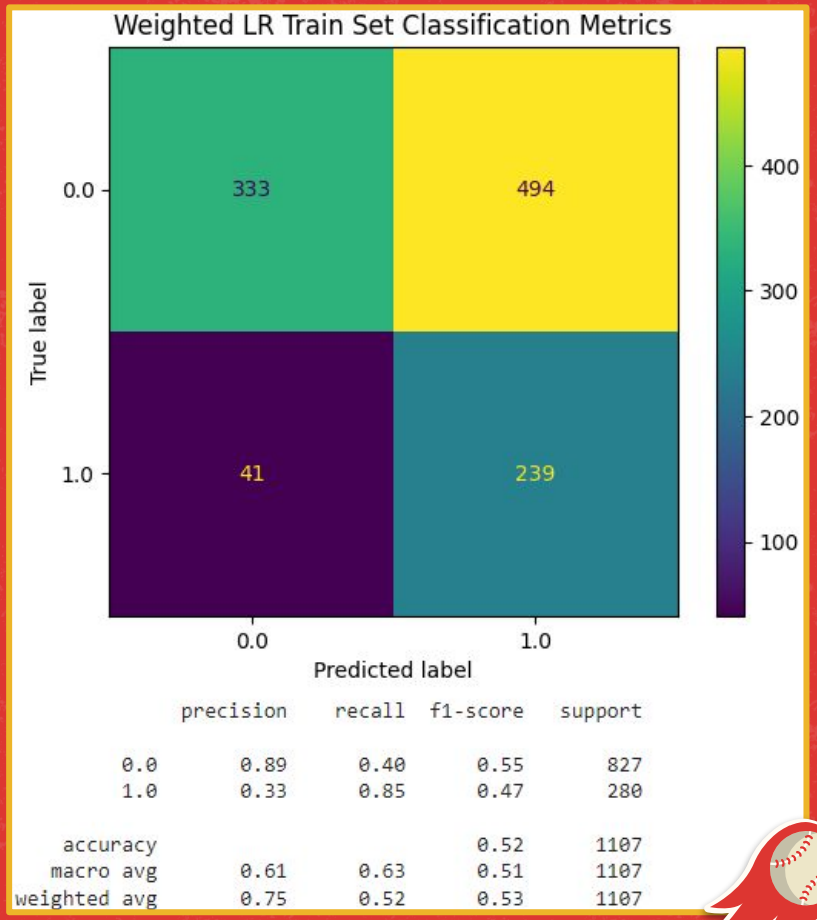
- Slider (SL)
- Sweeper (ST)
- Slurve (SV)
- Knuckle (KN)


Feature Importances
Age
Season
Intentional Walk Release Position X-dimension avg.
Fastball Release Position Y-dimension avg.
Curveball Release Position X-dimension avg.
Fastball Release Position Z-dimension avg.
Slow Curve Release Speed avg.
Fastball Release Speed avg.
Sweeper Velocity Y-dimension avg.
Changeup Velocity Z-dimension avg.
Fastball Velocity Y-dimension avg.
Sweeper Release Speed avg.
Throws
Slow Curve Velocity Y-dimension avg.
Pick-off Release Position X-dimension avg.
Sweeper Release Position Z-dimension avg.
Sweeper Release Position Y-dimension avg.
Knuckleball Release Position Y-dimension avg.
Cutter Release Position Y-dimension avg.
Knucklecurve Release Position Z-dimension avg.
4-seam Fastball Release Position Y-dimension avg.
Split-finger Release Position Z-dimension avg.
feature
−0.2
−0.1
0.0
0.1
0.2
0.3
importance

Best Model

Logistic Regression Model

Adjusted class weights



	precision	recall	f1-score	support
0.0	0.89	0.40	0.55	827
1.0	0.33	0.85	0.47	280
accuracy			0.52	1107
macro avg	0.61	0.63	0.51	1107
weighted avg	0.75	0.52	0.53	1107

Next Steps

Collect more data!

- Advanced pitching data prior 2008 (velocity, rotation, etc.)
- Video analysis - pitching mechanics
- Biological data (height, weight, diet, genetics, etc.)
- International leagues, minors, spring training, bullpen, college, etc.
- Other factors
  - Days rest, temperature, humidity, altitude, etc.

Questions?