

Predicting Tommy John Surgery

John Streaman





Business Understanding

Collect data on MLB pitchers to predict if they will need TJ 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 in 2022, 5 Year/ \$75 Million

Fees:

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

February

- **2 Bullpen sessions - 'arm fatigue'**
- **Shut down, MRI**
- **Shoulder strain, several weeks off**



Kodai Senga injury update: Mets' ace shut down for three weeks with shoulder strain, will miss Opening Day

Senga was originally slated to be the Mets' Opening Day starter



By [Matt Snyder](#) & [R.J. Anderson](#) Feb 25, 2024 at 9:58 am ET • 1 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

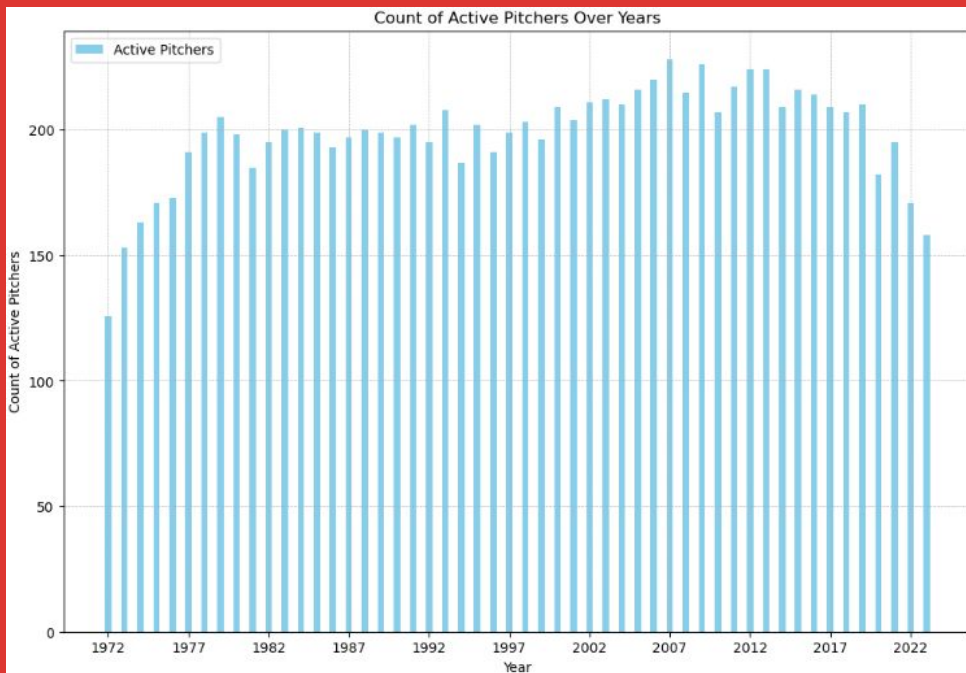
Historic Data

1972 - 2023

Starting Pitchers only

Minimum 320 career innings pitched

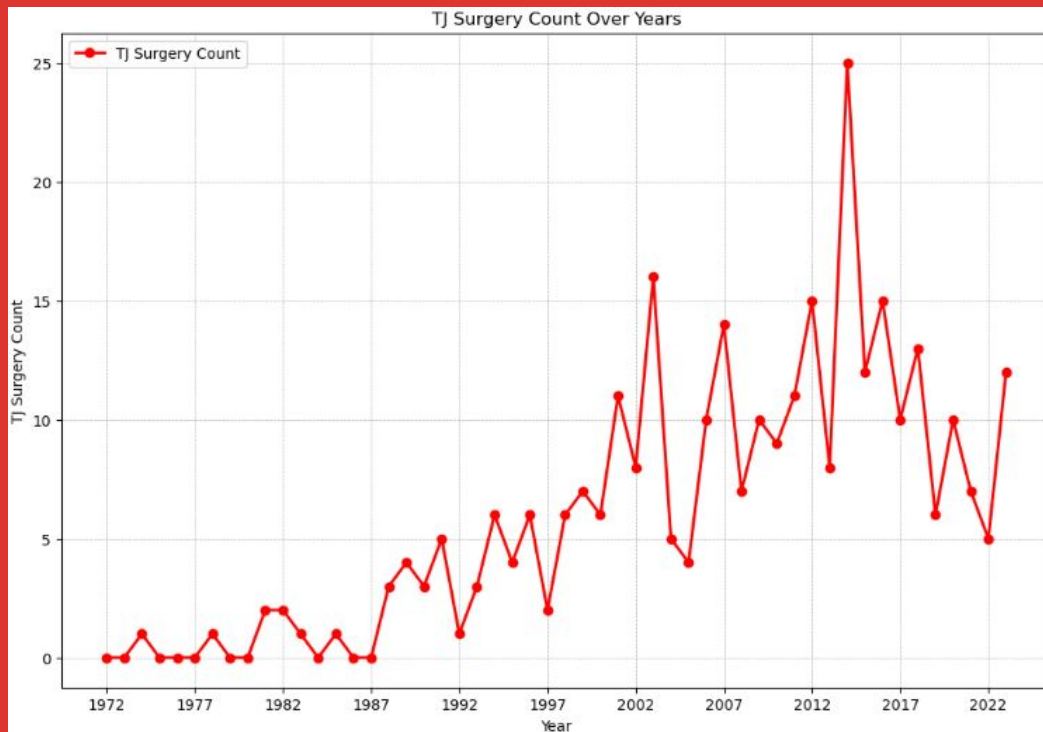
- Early 2000's
- Dramatic rise in number of pitchers



Historic Data cont.

TJ Surgery count

Dramatic spike 2003, 2014



Advanced Pitching Metrics

Seasonal data limited

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

Grouped data by season

- Season total count by pitch type
- Avg. speed, velocity, & release position
- Age, season, Throws (L/R)

Target - Surgery

- Binary (1.0 - yes, 0.0 - no)

Pitch types

Fastball (FB)

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

Off-speed (OS)

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

Breaking ball (BB)

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

Slider (SB)

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

Other (OT)

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

Best Model (thus far)

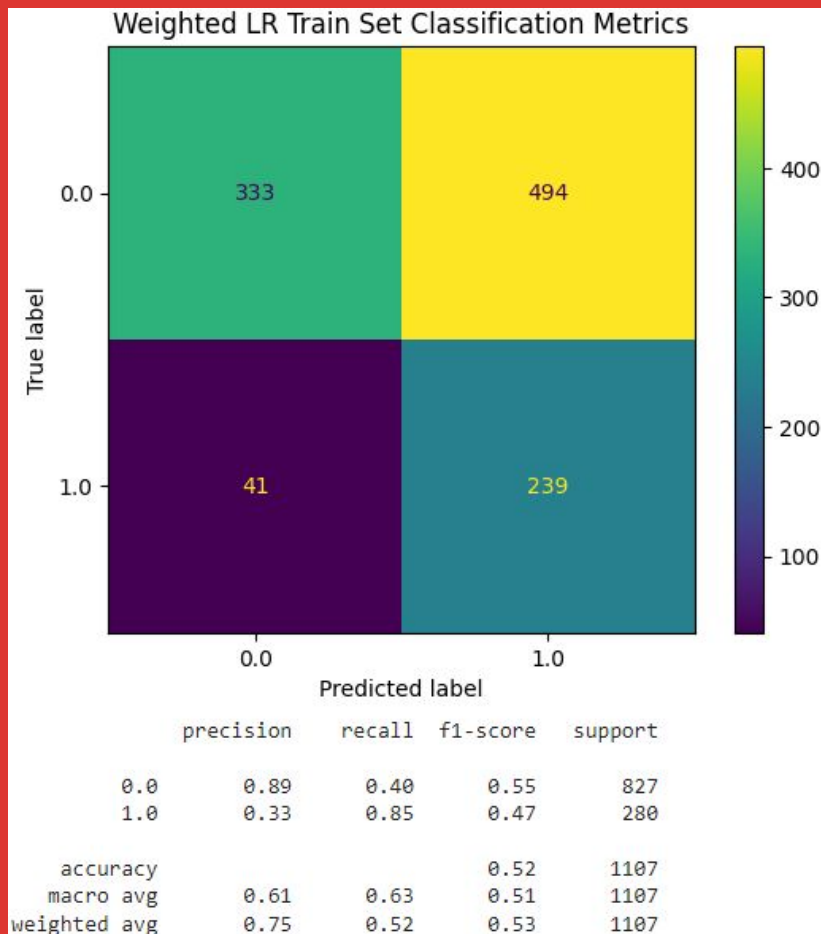
Logistic Regression Model

Standard Scaler

Changed class weights.

Score set to recall (medical)

Good recall score for predicting Surgery



Next best model

Decision Tree Classifier

Criterion = gini

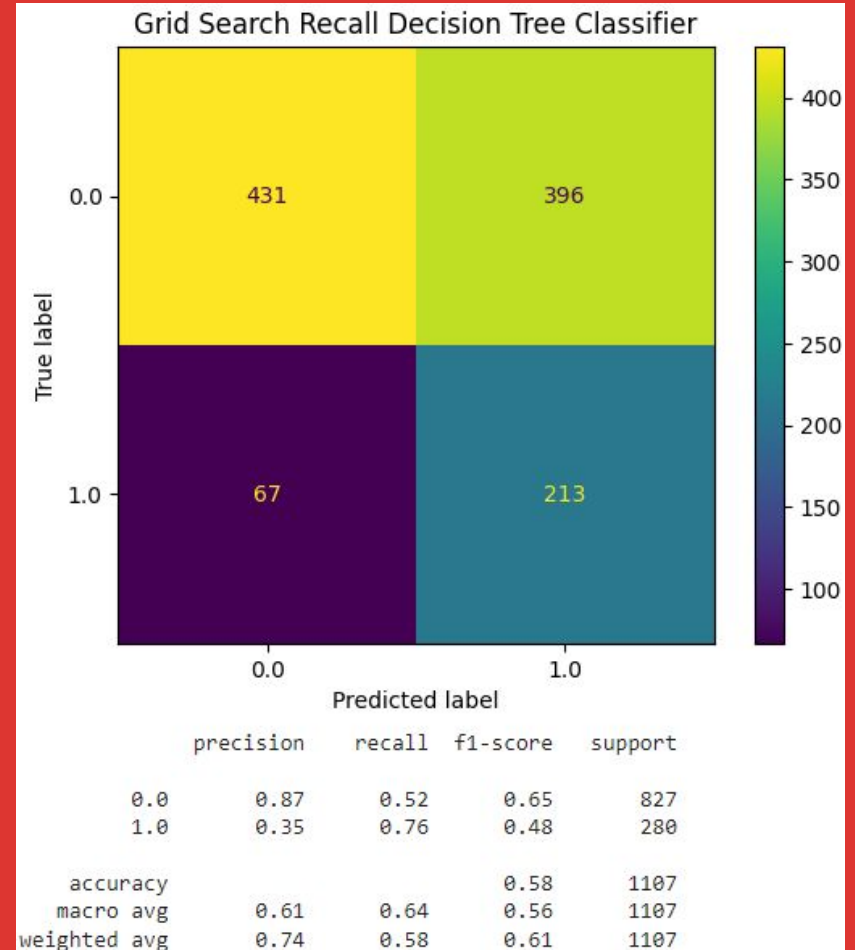
Class weights balanced

Max depth = 5

Min. samples leaf = 4

Score set to recall (medical)

Good recall score for predicting
Surgery



Next Steps

Collect Data from earlier seasons for more comprehensive knowledge

Analyze with video to better understand pitching mechanics

Gather data on other factors (height, weight, diet, weather, humidity, rest, etc.)

Collect data from minor leagues, international leagues, college, high school, etc.

The End.