MLB Multi-Out Play Mechanics

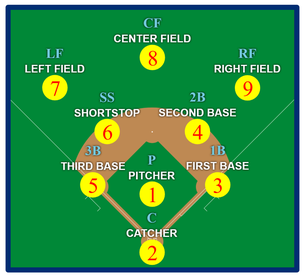
Cameron Perkins

Data Mining – Dr. Poonam Kumari

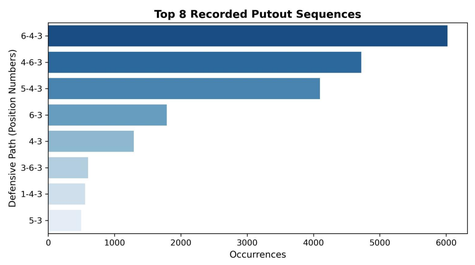
Bridgewater State University

Abstract

Multi-out plays (double and triple plays) are high-leverage defensive events in Major League Baseball that can rapidly shift game win expectancy. Using data from roughly 24,000 multi-out plays, this project applies unsupervised machine learning and association rule mining to uncover how pre-pitch context dictates defensive execution and to isolate extreme anomalies.



Baseball Positions by Number



Top 8 putout sequences on multi-out plays

Methods

Dataset & Preprocessing: ~24,000 MLB plays from regular and postseason games since November 2019, each with 119 pieces of data recorded on them.

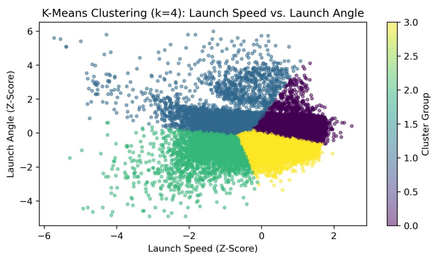
K-Means Clustering: Segmented batted-ball kinematics (Launch Speed, Launch Angle, Spray Angle, Hit Distance) to find distinct trajectory profiles (grounders vs. pop-fly outs).

Apriori Association Rule Mining: Mining frequent item sets across pitch type, count, matchup, and infield alignment, filtered to find and predict common put-out chains.

Isolation Forest & Sensitivity Analysis: Flagged the top 1% outliers (contamination = 0.01) in 4D space to find the weirdest and most unusual plays. A video clip of each play was also obtained via statcast.

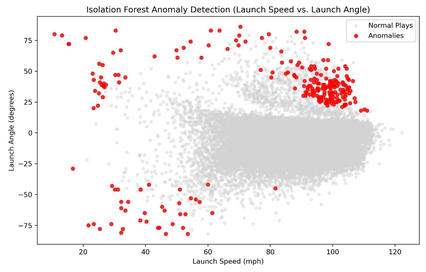
Results

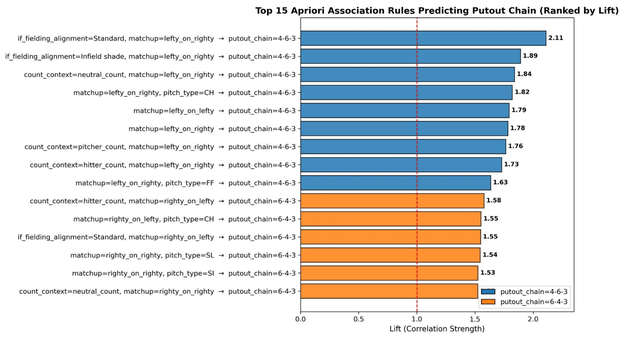
- **Matchup & Turn Direction:** Left-handed batters correlate strongly with 4-6-3 turns (lift - 2.11). Right-handed batters correlate with 6-4-3 turns (lift - 1.58).
- **Alignment & Pitch Type:** Using an Infield shade drops 4-6-3 confidence against lefties from 40.0% to 35.9% because middle infielders convert more balls into 6-3 or 3-6-3 plays. Changeups tend to make more 6-4-3 turns in Righty on Lefty matchups, while sliders and sinkers do in Righty-on-Righty matchups.
- **Outlier Detection:** The Isolation Forest (with a contamination of 0.01) flagged plays that have multiple extreme/abnormal metrics at once. The top 1% abnormal plays are mostly popped-up bunts near home plate and deep fly balls with baserunning outs.



K-Means clustering with 4 centroids, using 4 kinematic variables (Launch Speed, Launch Angle, Spray Angle, and Hit Distance) visualized in 2D via Launch Speed by Launch Angle

A scatterplot highlighting all plays recognized as an anomaly by the Isolation Forest (top 1%)





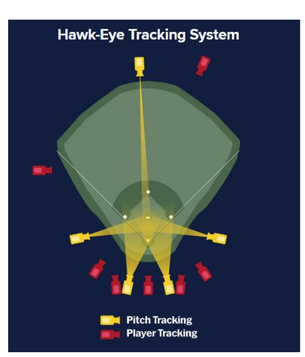
Tracking

Automated Video Retrieval: A function to match flagged anomalies to the MLB Stats API that automates finding broadcast video clips for manual review. Written in Python.



The play with the highest anomaly score: Geraldo Perdomo bunt pops into a double play in foul territory, catcher J.T. Realmuto to second baseman Bryson Stott. May 2nd, 2025
Arizona Diamondbacks @ Philadelphia Phillies

The Hawk-Eye Tracking System: A setup of 12 optical tracking cameras, implemented in all MLB stadiums to power Statcast analytics and the new ABS system as of 2026



Acknowledgements

Thanks to Dr. Poonam Kumari and the Bridgewater State University Computer Science Department for guidance on this project. Data collected from MLB Baseball Savant (Statcast) and the MLB Stats API.