Should I Stay or Should I Go

Billy Fryer

NC STATE UNIVERSITY

Department of Statistics

Background and Introduction

In baseball, every run counts and every decision matters. Even decisions such as attempting to advance 2 bases rather than just 1 could be the difference in the game and ultimately making the playoffs. For a ball hit to the outfield with at least one runner on first or second, I wanted to create a model to evaluate these decisions - whether the runner should stay after advancing one base or go and attempt to advance two bases (first to third or second to home). I then combined these probabilities with run expectancies to evaluate whether the potential advantage of being in a better future game state is worth the risk of being thrown out.

Four Factors

1) Distance From Base Runner to Target Base

2) Distance From Ball to Target Base

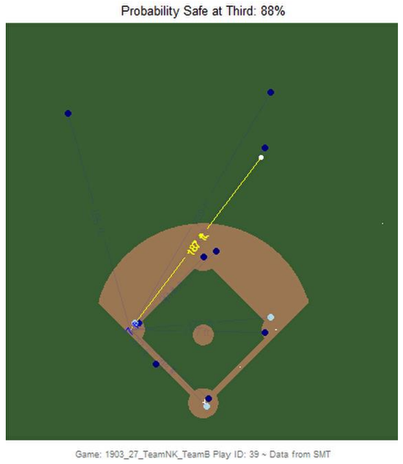
3) Base Runner Speed

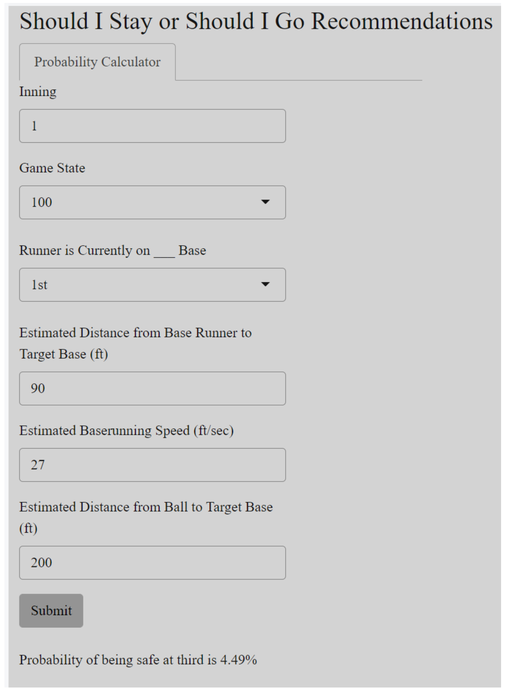
4) Game Situation Information

Modeling

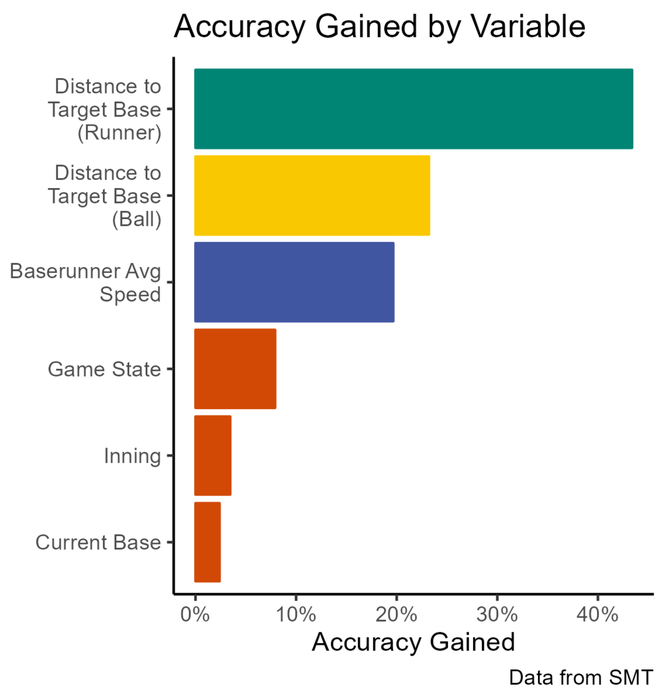
I tested 4 different types of classification models (logistic, naïve bayes, random forest, xgboost) and selected the xgboost model with 76.3% test set accuracy and a logloss of 0.290046 to create the probabilities for advancing 2 bases.

Scan the QR code above to view the play below with animation and test out the Shiny App!





In the play to the left, the ball is far from target base and runner is very close to target base → High Probability (88%)



Run Expectancy

The Stay/Go probabilities are great but by themselves are very hard to interpret. They need to be converted to a more familiar unit. Using the ideas from Marchi, Albert & Baumer, I converted the Stay/Go probabilities to run expectancies using the Change in Run Expectancy Formula below and the Run Expectancies from the 2015 AA Season. If the Change in Run Expectancy Formula results to a negative number, the number of expected runs would be less if the runner chose to attempt to advance, so they should just stay at the intermediate base. The opposite is true for a positive result.

Change in Run Expectancy Formula

RE(Success) x Pr(Success) + RS(Success) + RE(Failure) x Pr(Failure) + RS(Failure) – [RE(No Attempt) + RS(No Attempt)]

Where: RE() is the run expectancy of a game state
Pr() is the probability produced by the model
RS() is the runs scored from the play

Conclusions/Future Work

- This model can be used to evaluate if 3rd base coaches are too conservative in advancing runners
- Another use would be determining which prospects to call up in September as pinch runners
- Convert of the Stay/Go probabilities to the Run Expectancies in Shiny App
- Incorporate arm strength metric into model

Citation

Marchi, M., Albert, J., & Baumer, B. S. (2019). Chapter 5: Value of Plays Using Run Expectancy. In *Analyzing Baseball Data with R* (Second, pp. 111–136). essay, Taylor & Francis Group, LLC.

Thanks to SMT and UConn Sports Analytics Symposium for the opportunity to work with this data!