MLB Machine Learning Analytics: Pitch Prediction Model
2017 Spring Milestones

Alex Ahn

This document is an overview of the milestones to be accomplished for the Pitcher's Pitch Type Prediction model on a given timeline.

Milestones

<u>Feb 14</u>:

1. Completion of data ETL (Extract, Transform, Load) pipeline of datasets
   a. Revisions to the samples extraction will be completed

2. Validation of the dataset
   a. Sample dataset (Clayton Kershaw) along with instance dates will be sent out for validation

<u>Feb 21</u>:

1. Model Training Code Implementation: given datasets (csv files), code that trains each pitcher's prediction model will be implemented with following algorithms
   a. Neural Network (Multi Layer Perceptrons)
   b. Random Forest
   c. SVM (with multiple kernels)

<u>Feb 28</u>:

1. Model Selection Automation Process
   a. Hyper parameter tuning for Neural Network
   b. Hyper parameter tuning for Random Forest
   c. Hyper parameter tuning for SVM

<u>Mar 7</u>:

1. Evaluation: scores and results
   a. Baseline model tested and evaluated for each pitcher (scores recorded)

2. Discussions for adding new features
   a. Batter's statistics
      i. Home Runs history
      ii. Batting Average
      iii. WOBA+

Mar 14:

1. Feature Engineering
   a. Adding new features

2. Evaluation: scores and results
   a. Revised model tested and evaluated for each pitcher (scores recorded)

~ April:

1. Feature Engineering (II)
   a. More feature tunings, additions, or eliminations

2. Evaluation: scores and results
   b. Revised model tested and evaluated for each pitcher (scores recorded)

April ~ May:

1. Paper Write-up
   a. From Abstract to Conclusion
   b. Summarizing results

May ~ end:

1. Model / Paper Review

2. Discussions on future usage of the model / paper