Appendix B: Contributions

Mehmetali Kulunyar

* Data scraping/filtering of MLB Pitchf/x data
* Created visuals for Random Forest, Decision Trees, ‘Net Gain’ analysis
* Feature Engineering: Normalization, Dummy Variables, N-1, n-2, n-3 pitch type, n-1, n-2, n-3 pitch result, Cumulative Pitch Count, Game Type Description
* Decision Tree model analysis for different parameters (accuracy, log-loss)

Shangying Jiang

* Data preparation: converted string objects into numerical objects depending on features’ properties
* Filled missing values using various strategies (mean, most frequent value)
* Created indicator features for some features with missing values, plus ‘scoring position’
* Built prototypes for logistic regression, decision tree and random forest
* Plotted feature importances for decision tree and random forest
* Test performance of models using different parameter settings

Brenton Arnaboldi

* Created features for ‘batter position’ and ‘pitch type percentage’ from the previous season
* Imported batting and pitching features from Baseball Prospectus
* Formulated the concept of ‘Net Gain’
* Lead writer for the proposal and status report

Dave Hamilton