**How to Predict a Playoff Bound Baseball Team**

The goal of this project is to build an algorithm that will predict the winner in each of the six Major League Baseball (MLB) divisions based on their statistical measures of a team’s performance. Winning an MLB division is important because it means a playoff matchup. The success of any baseball team is based on two things:

* Winning baseball games
* Making a profit

In order to win a division, a team would have to win enough baseball games. The more successful a team is, the more fan involvement there will be. People will buy tickets for the games and also fans will support the team with merchandise sales. Making the playoffs by winning a division guarantees more home field baseball games and in turn, more ticket sales for the year.

For these reasons, people who will be interested in this project would be scouts and general managers since the developed algorithm will enable them to assess talent in a more accurate fashion for what will benefit the team to create more wins.

I will be using the Kaggle Baseball databank that has data on baseball teams from 1871 to 2015. I will be using approximately 26 features, such as home runs, strikeouts, division wins, and many others. I will formulate the problem as a binary classification problem and I will utilize the scikit-learn library to build the predictive models. Candidate models include logistic regression, random forests, and support vector machines.

The deliverables of this project will be the code in the form of a Jupyter notebook, a report describing the work I did in full detail, and a blog post.