Baseball Case Study

Problem Statement:

This dataset utilizes data from 2014 Major League Baseball seasons in order to develop an algorithm that predicts the number of wins for a given team in the 2015 season based on several different indicators of success. There are 16 different features that will be used as the inputs to the machine learning and the output will be a value that represents the number of wins.

**-- Input features:** Runs, At Bats, Hits, Doubles, Triples, Homeruns, Walks, Strikeouts, Stolen Bases, Runs Allowed, Earned Runs, Earned Run Average (ERA), Shutouts, Saves, Complete Games and Errors

**-- Output:** Number of predicted wins (W)

To understand the columns meaning, follow the link given below to understand the baseball statistics: https://en.wikipedia.org/wiki/Baseball\_statistics

**For downloading the dataset, use the link given below.**

Download Files:

* <https://github.com/dsrscientist/Data-Science-ML-Capstone-Projects/blob/master/baseball.csv>

**Statistics (https://en.wikipedia.org/wiki/Baseball\_statistics)**

* R – [Runs](https://en.wikipedia.org/wiki/Run_(baseball)) scored: times reached home plate legally and safely
* AB – [At bat](https://en.wikipedia.org/wiki/At_bat): plate appearances, not including bases on balls, being hit by pitch, sacrifices, interference, or obstruction
* H – [Hit](https://en.wikipedia.org/wiki/Hit_(baseball)): reaching base because of a batted, fair ball without error by the defense
* 2B – [Double](https://en.wikipedia.org/wiki/Double_(baseball)): hits on which the batter reaches second base safely without the contribution of a fielding error
* 3B – [Triple](https://en.wikipedia.org/wiki/Triple_(baseball)): hits on which the batter reaches third base safely without the contribution of a fielding error
* HR – [Home runs](https://en.wikipedia.org/wiki/Home_run): hits on which the batter successfully touched all four bases, without the contribution of a fielding error
* BB – [Base on balls](https://en.wikipedia.org/wiki/Base_on_balls) (also called a "walk"): hitter not swinging at four pitches called out of the strike zone and awarded first base.
* K (or SO) – [Strikeout](https://en.wikipedia.org/wiki/Strikeout): number of batters who received strike three
* SB – [Stolen base](https://en.wikipedia.org/wiki/Stolen_base): number of bases advanced by the runner while the ball is in the possession of the defense
* HR (or HRA) – [Home runs allowed](https://en.wikipedia.org/wiki/Home_runs_allowed): total home runs allowed
* ER – [Earned run](https://en.wikipedia.org/wiki/Earned_run): number of runs that did not occur as a result of errors or passed balls
* ERA – [Earned run average](https://en.wikipedia.org/wiki/Earned_run_average): total number of earned runs (see "ER" above), multiplied by 9, divided by innings pitched
* SHO – [Shutout](https://en.wikipedia.org/wiki/Shutout_(baseball)): number of complete games pitched with no runs allowed
* SV – [Save](https://en.wikipedia.org/wiki/Save_(baseball)): number of games where the pitcher enters a game led by the pitcher's team, finishes the game without surrendering the lead, is not the winning pitcher, and either (a) the lead was three runs or fewer when the pitcher entered the game; (b) the potential tying run was on base, at bat, or on deck; or (c) the pitcher pitched three or more innings
* CG – [Complete game](https://en.wikipedia.org/wiki/Complete_game): number of games where player was the only pitcher for their team
* E – [Errors](https://en.wikipedia.org/wiki/Error_(baseball_statistics)): number of times a fielder fails to make a play he should have made with common effort, and the offense benefits as a result
* W – [Win](https://en.wikipedia.org/wiki/Win_(baseball)): number of games where pitcher was pitching while their team took the lead and went on to win, also the starter needs to pitch at least 5 innings of work