Sports Betting Data Science Project

June 22nd, 2022

Sam LaFell

# About the Project:

This project was born out of a conversation with a friend of mine. He talks a lot about sports betting, and so did a former coworker of mine. After hearing people talk about how much money you can make (and lose) while doing this, I took an interest in it.

Then, I started thinking, it’s really just odds. Which is why so many sports outlets use “odds” when they talk about the payouts when you put down certain bets.

So, we can easily use data science. Also, there is a lot of data out there to be able to use and build a model on to predict.

## Initial Objective:

Predict over/unders in baseball. For example, if a game has an over/under of 9.5, it means if the two teams combine for 9 runs, you would win by taking the under, and if they combine for 10+ runs, you would win by taking the over.

We are going to use over/unders because you can put down like $20 on each one, and walk away with $30-40, or of course lose your $20. So, it’s minimal loss once we get into actually testing the model in the real world.

Also, there are so many baseball games, it will be easier to backtest and deploy the models.

## Things we won’t do:

Deploy the algorithm to a point where it is making bets for us. This is V1 of the project, and the goal is to produce an algo that provides a decision, but we must actually make the bets. This will just be safer for our wallets.

## Learning Objectives:

* Language:
  + Python
* Code
  + A focus on good coding practices here. Should be a good project to be able to test out functions, classes, data structures all in Python. A focus on repeatable code.
  + Unit Testing of Classes/Functions to make sure we are obtaining expected results, another thing that gets looked over a lot in DS training
* Version Control
  + GIT! Working in a small team, GIT will be our go to for version controlling. This will help us get a better grasp of how to work with it
* Project Management
  + Either Tello or Asana
* Algorithms
  + XGBoost
  + Neural Networks
  + Decision Trees and Random Forests
* Data Science Approaches
  + Simulation/Risk Analysis
  + Web Scraping
  + Possibly Game Theory

# Data Resources:

* https://www.seanlahman.com/baseball-archive/statistics
* Baseball Reference
* FanGraphs
* Baseball Data Science
* MLB website
* <https://www.retrosheet.org/>

# Proposed Timeline:

**July:** Begin project, solidify team, break down larger goals into smaller pieces

**August-November:** Coding work, including weekly or biweekly meetings on Tuesdays or Thursdays. These can be code reviews, brainstorm for future steps, or anything. Max length, 90 minutes.

**December**: No work

**January:** Meeting to kick off the new year, and evaluating where we are in the model process and how much is left to be done

**February:** Finalizing the model, including back tests

**March:** Hoping the model is ready for the start of the upcoming season