INFSCI1091 Moneyball 2.0: Winning in Sports with Data

Spring 2018
Class meetings: Tue/Thu 11am-12:15pm

Konstantinos Pelechrinis
Office hours: TBA
kpele@pitt.edu
Contact by e-mail or during office hours

Course Description Data and analytics have been part of the sports industry from as early as the 1870s, when the first boxscore in baseball was recorded. However, it is only recently that advanced data mining and machine learning techniques have been utilized for facilitating the operations of sports franchises. While part of the reason is related with the ability to collect more fine-grained data, an equally important factor for this turn to analytics is the huge success and competitive advantage that early adopters of investment in analytics enjoyed (popularized by the best-seller ``Moneyball'' that described the success that Oakland Athletics had with analytics). Draft selection, game-day decision making and player evaluation are just a few of the applications where sports analytics play a crucial role today. Apart from the sports clubs, other stakeholders in the industry (e.g., the leagues' offices, media, etc.) invest in analytics. The leagues increasingly rely on data to decide on potential rule changes. In this course, we will introduce data science concepts for sports analytics. Students will learn concepts related to data collection, data analysis and modeling as well as data visualization.

Prerequisites: STAT 200
Recommended prerequisites: STAT 1100, CS 8 or equivalents

Reference Textbook (optional): W. Winston, “Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball and Football”, Princeton University Press, 2009.

Course Requirements & Grading:

The evaluation will be based on the following components:

1) Homework: 30%
2) Project: 20%
3) Midterm: 25%
4) Final: 25%

There will also be a **data competition for extra credit**. The maximum extra credit for a student will be 10%. The formula for the total extra points given to student i is:

$$EC_i = 10\frac{N\text{-}r_i}{N-1}$$

where r_i is the rank of student i at the competition and N is the total number of students that participated in the competition

Tentative Schedule

Week	Topics covered
Week 1	Where are analytics used in sports? What is the current state-of-the art? **Data collection**: API use, web scraping
Week 2	**Empirical probability and statistical tests** **concepts**: two-point conversion in NFL, fourth-down decision in the NFL, end-game basketball strategy, do officials fix NBA games?, is the hot-hand real?
Week 3	**Introduction to Monte Carlo simulations** **Introduction to resampling**: Bootstrap
Week 4	**Introduction to regression**: Bradley-Terry model, Oliver's four factors, in-game win probability **Evaluating probability models**: Brier score, probability validation curves
Week 5	**Team rating systems**: Elo rating, regression-based rating, network-based rating **Player rating systems:** QBR, adjusted plus/minus, VORP, wins contributed
Week 6	**Regularization:** Preventing overfitting **Simulating a tournament:** Combining discrete event simulations and team ratings
Week 7	Midterm
Week 8	**Strength of Schedule** **NFL point values per play**
Week 9	**Draft analytics**
Week 10	**Spatial sports data:** SportVU, Statcast, NextGen **Floor spacing and shot selection in the NBA**
Week 11	**Introduction to Game Theory:** (mixed) strategies, Nash Equilibrium **Case studies:** passing vs rushing in NFL, penalty kicks in soccer
Week 12	**Clustering:** grouping players and teams
Week 13	**Networks:** passing networks, substitution networks **Network embedding:** evaluating lineups
Week 14	Final

University Policies:

Collaboration Policy

Miderm: No collaboration

Final: No collaboration

Homework: No collaboration. However, discussion is encouraged, but stay away from one another answers. When discussing a problem, avoid taking notes. If you filter the discussion through your own memory, you are very unlikely to submit answers that "accidentally" indicates you collaborated. This is for your own protection.

Material Covered

You are responsible for all material covered in lectures, as well as, reading and homework assigned.

Academic Integrity

Students in this course will be expected to comply with the University of Pittsburgh's Policy on Academic Integrity. Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

G-grade Policy

No G-grade will be granted and students must finish their work load during the semester.

Disability Services

If you have a disability that requires special testing accommodations or other classroom modifications, you need to notify both the instructor and Disability Resources and Services no later than the second week of the term. You may be asked to provide documentation of your disability to determine the appropriateness of accommodations. To notify Disability Resources and Services, call (412) 648-7890 (Voice or TTD) to schedule an appointment. The Disability Resources and Services office is located in 140 William Pitt Union on the Oakland campus.

Copyright Notice

Course materials may be protected by copyright. United States copyright law, 17 USC section 101, et seq., in addition to University policy and procedures, prohibit unauthorized duplication or retransmission of course materials. See Library of Congress Copyright Office and the University Copyright Policy.

Statement on Classroom Recording

To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the

instructor, and any such recording properly approved in advance can be used solely for the student's own private use.