

Kenneth P. Dietrich School of Arts and Sciences
College in High School

2023-2024

Decision-Making in Sports
INFSCI 0530--3 Credits

Description: Data and analytics have been part of the sports industry from as early as the 1870s, when the first box score 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 get introduced concepts related to data collection, data quality, data analysis and modeling as well as data visualization.

Prerequisites: STAT 0200, STAT 1100, CS 0012 or equivalents are recommended.

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

Learning outcomes:

Upon successful completion of the course students will be able to:

- Understand how to interpret and use probabilities both in the field of sports decision making, as well as in general
- Become familiar with basic data analysis techniques, such as linear and logistic regression
- Obtain a gentle introduction to Monte Carlo simulations
- Understand the notion of overfitting and how you can avoid it
- Understand what comprises a useful visualization and why effective communication is crucial

Course Requirements & Grading:

The evaluation will be based on the following components:

- Homework 20%
 - Project 30%
 - Midterm 25%
 - Final 25%
-

Kenneth P. Dietrich School of Arts and Sciences
College in High School

Course Topics:

1. Data collection, quality and ethical concerns

- where are analytics used in sports?
- what is the current state-of-the art?

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?

3. Bayesian player evaluation

4. Sports Science and Analytics

5. Introduction to regression

- Bradley-Terry model
- Oliver's four factors
- in-game win probability

6. Evaluating probability models

- Brier score
- probability validation curves

7. Team rating systems

- Elo rating
- regression-based rating
- network-based rating

8. Player rating systems

- QBR
- adjusted plus/minus

9. What is overfitting? How to avoid it

- regularization and its Bayesian interpretation

10. Occam's razor and model selection

11. Introduction to Monte Carlo simulations

- simulating a tournament

12. What is resampling?

13. Sports and behavioral economics

- decision making and risk adversity

14. Natural experiments

- the case of corner three shots

15. Why are visualizations important?

- principles of visualizations

16. Introduction to Game Theory

- (mixed) strategies
- Nash Equilibrium

17. Case studies

- passing vs rushing in NFL
- penalty kicks in soccer
- defending the corner three in the NBA

18. Unsupervised learning clustering

- grouping players and teams

Kenneth P. Dietrich School of Arts and Sciences

College in High School

Course Policies:

Collaboration Policy

- *Midterm:* 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" indicate 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.

Related educational materials:

High school teachers are encouraged to use the following educational modules while teaching the course.

The modules can be downloaded from the SADET project website:

http://www.sis.pitt.edu/lersais/research/sadet/module_menu.html

| Relevant SADET modules |
|---|
| Module B.2: Basic Statistical Concepts |
| Module B.3: The Notion of Uncertainty |
| Module B.1: Evidence-based Thinking Module |
| Module D.4: Data Mining techniques |
| Module D.6: Privacy-aware Data Mining |
| Module D.8: Email Classification and Spam Detection |

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](#).

Academic Integrity: All College in High School teachers, students, and their parents/guardians are required to review and be familiar with the University of Pittsburgh's Academic Integrity Policy located online at <https://www.as.pitt.edu/faculty/policies-and-procedures/academic-integrity-code>.

Grades: Grade criteria in the high school course may differ slightly from University of Pittsburgh standards. A CHS student could receive two course grades: one for high school and one for the University transcript. In most cases the grades are the same. These grading standards are explained at the beginning of each course.

Kenneth P. Dietrich School of Arts and Sciences
College in High School

Transfer Credit: University of Pittsburgh grades earned in CHS courses appear on an official University of Pittsburgh transcript, and the course credits are likely to be eligible for transfer to other colleges and universities. Students are encouraged to contact potential colleges and universities in advance to ensure their CHS credits would be accepted. If students decide to attend any University of Pittsburgh campuses, the University of Pittsburgh grade earned in the course will count toward the student grade point average at the University. At the University of Pittsburgh, the CHS course supersedes any equivalent AP credit.

Drops and Withdrawals: Students should monitor progress in a course. CHS teacher can obtain a Course Drop/Withdrawal Request form from the CHS office or Aspire. The form must be completed by the student, teacher and parent/guardian and returned to teacher by deadlines listed. Dropping and withdrawing from the CHS course has no effect on enrollment in the high school credits for the course.