

INFSCI1091 Moneyball 2.0: Winning in Sports with Data

Project Description

This will be a semester-wide project that you must complete by the end of the semester. It is an open-ended project, i.e., even though suggestions will be provided you can choose a topic of your interest related to sports and analytics. I will then provide recommendations and feedback to define the specific problem to be solved and the deliverables. Projects can either be individual or groups of two students. In a group of two students, both students will get the same grade. So, there are trade-offs in having group-versus-individual projects that I will let to you consider.

Each proposed project needs to include the following subtasks:

- **Problem motivation and review:** Why the project you are proposing is important in your opinion? What is the current state-of-the-art? Why current solutions/answers are not satisfying?
- **Data collection:** You will need to identify and collect the appropriate data to complete the proposed project. Why these data are appropriate for answering your questions?
- **Descriptive analysis:** You will need to perform an exploratory analysis on the dataset you collected. How do the data look like? What is their distribution? Are there missing data?
- **In-depth analysis/modeling:** At this part of the project you will need to delve into a deep analysis of the data to answer your question. This might require appropriate models or simply in-depth analysis of the data. In case of modeling, appropriate evaluation of the models need to be performed. In case of in-depth analysis convincing arguments based on the results need to be given for the conclusions drawn.
- **Visualizations:** Concrete and clear visualizations of the results need to be provided so that even a non-expert in data analysis can grasp the conclusions and results.
- **Report & code:** You must deliver all the above within a report and provide any source code you generated during the project.

Timeline

During the first 4 weeks of the course you must submit to me a one-page description of your project. This one-page will describe the above steps at a high level, with emphasis on the motivation for the problem and why you think it is important/interesting. Once we agree upon the exact question(s) of the project you will be able to start working on it. During the last week of classes, we will have project presentations (either in a classroom setting or as a poster session depending on the number of projects). The final report will be due the last day of classes.

Project suggestions

The following are simple suggestions and you are not required to choose any of these projects.

1. Does the bye week in NFL, or the back-to-back games in NBA, impact the chances of winning a game?
2. What is the probability that Ben Roethlisberger (or any player for that matter) will be a first ballot hall of famer?
3. Are Golden State Warriors better than the 1990s Chicago Bulls?
4. Are there referee biases in the NBA (or any other league)?
5. How can you quantify the value of a pass in basketball (or any other sport)?
6. What is the value of a timeout? For example, how much is your win probability affected if you have all your 3 timeouts at the two-minute warning in the fourth quarter in NFL? Can timeouts stop momentum in basketball?
7. Did the change of the kickoff touchback line achieve its goal, i.e., eliminate kickoff returns and hard player collisions? If no, what change rules could potentially achieve this goal?