

2022 UConn Sports Analytics Symposium (UCSAS)

Saturday, October 8, 2022

<https://stats.org/events/ucsas2022/>

Back in-person and supported by the NSF in 2022, the UCSAS focuses specifically on students (graduate, undergraduate, and pre-college) who are interested in sports analytics. Organized by the UConn Statistical Data Science Lab of the Department of Statistics, UCSAS aims to: 1) showcase sports analytics to students at an accessible level; 2) train students in data analytics with application to sports data; and 3) foster collaboration between academic programs and the sports industry.

Keynote Presentations

- Tegan Bunsu Ashby, Senior Full-Stack Developer, Basketball Systems, Brooklyn Nets
“Software Engineering for Sports Analytics”
- David Bergman, Associate Professor, University of Connecticut
“Integration of Analytics Techniques for Algorithmic Sports Betting”

Panel Discussion: Sports Data Science Competitions

- Alison Lukan (moderator), Seattle Kraken Contributor and TV Analyst for Root Sports.
- Michael Lopez, Senior Director of Football Data and Analytics, National Football League.
- Brendan Kumagai, MS Student in Statistics, Simon Fraser University (2022 Big Data Bowl Winner).
- Megan Risdal, Product Manager, Kaggle.
- Asmae Toumi, Director of Analytics, PursueCare (2021 Big Data Bowl Winner).

SMT Data Challenge The newly added data challenge is sponsored by SportsMedia Technology (SMT), an industry leader in sports data collection and visualization. SMT has provided anonymized in-game player and ball location data for multiple minor league teams over multiple seasons. The goal of the data challenge is to analyze an aspect of player movement (e.g., baserunning, movement while fielding, backing up a play) within this spatiotemporal dataset. The submission deadline is Monday, August 1 at 11:59pm. Finalists selected by the data challenge judging committee will be invited to present their works in the poster session with travel support, and the winner will be announced at the closing ceremony.

Poster Session A virtual poster session is scheduled for 11:30–13:30. We invite submissions from all, especially students (pre-college, undergraduate, or graduate, with travel support), with interesting works on any topics of sports analytics. The submission deadline is 11:59 pm, Friday, September 30, 2022. A student poster award, decided by the Student Poster Award Committee, will be presented at the closing ceremony.

Training Workshops Six 50-minute workshops are offered in three concurrent tracks during 13:45–15:35. They provide trainings from jumpstart to advanced sports analytic skills.

| | Introductory Track | Intermediate Track | Advanced Track |
|-------------|------------------------|--------------------|--------------------------------|
| 13:45–14:35 | Introduction to R | Hockey Analytics | Web Scraping for Sports Data |
| 14:45–15:35 | Introduction to Python | Baseball Analytics | TensorFlow in Sports Analytics |

Registration The symposium is open to anyone with an interest in sports analytics. The registration fee is minimal (\$5 for students and \$10 for non-students) to get an accurate count for planning and logistics.

Sponsorship Depending on the contribution level, sponsors will be recognized through customary channels, including names of sub-events. Please contact Dr. Jun Yan (jun.yan@uconn.edu) for details.



SMT DATA CHALLENGE

For this data challenge, your goal is to analyze an aspect of player movement (e.g. baserunning, movement while fielding, backing up a play) for minor league baseball players. SMT has provided in-game player and ball location data for multiple teams over multiple seasons. This spatiotemporal information can fuel a thoughtful analysis to answer questions that are difficult or impossible to answer with manually collected data or subjective observation. Since this challenge provides the opportunity to work with previously unavailable player tracking data, your analysis should involve player motion; this includes any topic that uses player location data over time. Below are a few example topics.

| BASERUNNING

- How could you evaluate a batter's chance to advance to second base on a ball in play? Or, for a runner on first base, the chance to advance from first-to-third on a single or first-to-home on a double?
- What circumstances are most likely to induce pickoff throws? Given a situation involving pickoff throws, what baserunner behavior is most likely to result in a stolen base? A successful pickoff?
- How would you evaluate a baserunner's ability to read a ball in play?

| FIELDING

- When a ball in play reaches the outfield, what aspects of fielding are most important in preventing a baserunner from advancing, or a batter from taking extra bases?
- How would you evaluate a player's fielding ability in the context of judgment and risk-taking? For example, how would you compare a player who attempts and fails to make a difficult play (possibly leading to an error) to a player that does not attempt to make a play?
- How would you evaluate a fielder's ability to read a ball in play?
- What attributes are most important to fielding and assisting on infield groundouts? Double plays?

| GENERAL

- How could you estimate expected runs throughout the course of a play based on player and ball locations and use that to evaluate player baserunning and fielding?
- Which baserunning and fielding abilities are most predictive or most consistent from season to the next?

We emphasize that this list is not exhaustive, and participants should feel free to study an aspect of player movement that interests them.

| REGISTRATION

Participants can register a team of up to four people at the [SMT Data Challenge Registration Page](#)

| SUBMISSION REQUIREMENTS

Please submit

- A short paper on your study in PDF format (max: 3000 words)
- A GitHub repo link containing code files and .csv files with results

Submissions are due by Monday August 1, 11:59pm Eastern Daylight Time.

| JUDGING CRITERIA

A panel of judges from across academia and the sports industry will judge your submissions based on the following:

- How original is the analysis?
- How applicable is the analysis?
- How appropriate were the methods used?
- How well did you communicate your findings? This includes both written text and visualizations. How did the use of facts, data-supported narratives, anecdotes, visual aids, etc. buttress storytelling?

We will notify winners in early September.