

Data Science Week 2020

November 30th - December 4th

<https://sites.google.com/view/data-science-week-2020/home>

Monday 30th November 2020 1:30pm KT216

Prof. Bin Yu

Departments of Statistics, Electrical Engineering, Computer Sciences - University of California, Berkeley

Veridical data science:

the practice of responsible data analysis and decision making



Veridical data science aims at responsible, reliable, reproducible, and transparent data analysis and decision-making. Predictability, computability, and stability (PCS) are three core principles towards veridical data science. They embed the scientific principles of prediction and replication in data-driven decision making while recognizing the central role of computation. Based on these principles, the PCS framework consists of a workflow and documentation (in R Markdown or Jupyter Notebook) for the entire data science life cycle from problem formulation, data collection, data cleaning to modeling and data result interpretation and conclusions.

Employing the PCS framework in causal inference and analyzing data from clinical trial VIGOR, we developed staDISC for stable discovery of interpretable subgroups via calibration for precision medicine. The subgroups discovered by staDISC using the VIGOR data is validated to a good extent with the APPROVe study.

Virtual Presentation: <https://purdue.webex.com/meet/aselvite>

Data Science and Machine Learning Seminar Series || Data Science and COVID-19 Thematic Program 2020-2021

<https://sites.google.com/view/data-science-epidemiology> || <https://users.pfw.edu/aselvite/DSMLS/index.html>