BST 209: Collaborative projects
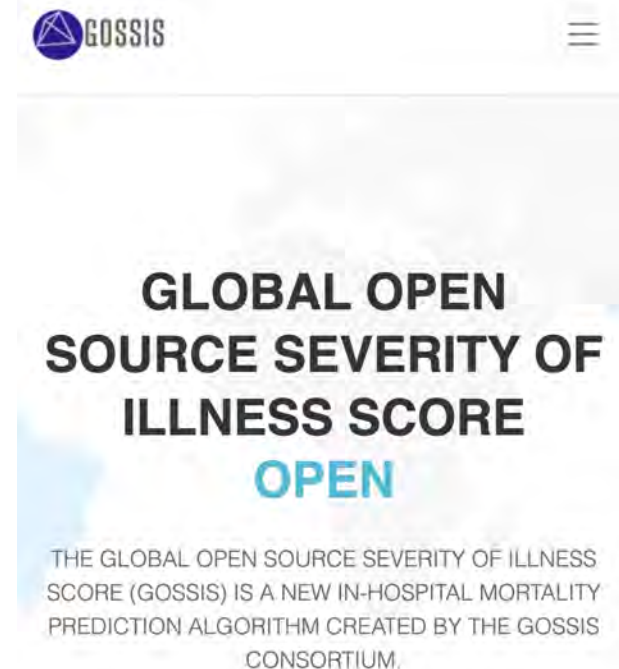
Wednesday, 6 July 2022

Schedule

- Introduction to ML (5-15 July)

- Team projects (18-21 July)

- Presentations (22 July)

Datasets

"GOSSIS" subset

- Variables collected in first 24 hours admission to ~200 ICUs across the United States.

- Curated CSV, derived from the eICU Collaborative Research Database (CRD)

- Labels (in-hospital mortality, age, sex, comorbidities)

GOSSIS

GLOBAL OPEN
SOURCE SEVERITY OF
ILLNESS SCORE
OPEN

THE GLOBAL OPEN SOURCE SEVERITY OF ILLNESS
SCORE (GOSSIS) IS A NEW IN-HOSPITAL MORTALITY
PREDICTION ALGORITHM CREATED BY THE GOSSIS
CONSORTIUM.

MIMIC-IV

- Publicly accessible EHR database

- >50,000 ICU stays between 2008-2018

- Modular (structured EHR, X-rays, waveforms, clinical text, echos...)

- Extensively used across education, research, and industry

PhysioNet

- Data sharing platform built and maintained by MIT

- Supports data use agreements and training

- Hosts both the GOSSIS and MIMIC-IV datasets

Team project

Presentation (Friday 22 July)

- Project of your choice!

- 5 minute talk (group, slides)

- Reproducible analysis (group, RMarkdown)

Suggestion

- ICUs often lack verified medical histories for incoming patients.

- Knowledge about chronic conditions can inform clinical decisions about patient care and ultimately improve patient outcomes.

- Using the GOSSIS dataset, determine whether a patient admitted to an ICU has been previously diagnosed with diabetes.

- **Previously a WIDS Kaggle competition, so lots of existing public work on this topic.**

Alternatives

- Using MIMIC-IV, predict hospital or ICU admission using emergency department data.

- Choose your own topic (with the caveat that time is very limited!)

Goals

Create and apply code

Become familiar with ML concepts

13

Build collaborators

Teams

Team	Students	Data Scientists
Team A	Nikola Stankovic Ben Albert	Xavier Borrat Adrien Carrel
Team B	Prakriti Gaba Lucas Chen Najim Lahrouchi	Li-Ching Chen Aubrey Tseng
Team C	Kandice Kapinos Nurit Katz	Jack Gallifant Bhav Jain
Team D	Claire Shappell Gregory McDermott	Sicheng Hao
Team E	Rosangela Hoshi Ye Zhang	Yugang Jia Yanran Li

Team	Students	Data Scientists
Team F	Shernaz Dossabhoy Marcelo Furtado Jennifer Cape	Po-Chih Kuo
Team G	Kendra Wulczyn Justine Ryu	Gloria Kwak
Team H	Junhan Zhao David Roach Eugenia Uche-Anya	Luis Nakayama Arnoud Petit
Team I	Annie Hsieh Julian Kleine-Borgmann Abdelsimar Omar	Ryan Wang Sulaiman Moukheiber
Team J	Brendan Striano Alexander Crawford	Yi-Ting Tsai Aqdas Kamal

Any questions?