Stanford
MEDICINE | Biomedical Data Science

DBDS Workshop in Biostatistics
Remote Access Only:
Contact kkanagaw@stanford.edu for Zoom dial-in details.

DATE:	September 17, 2020
TIME:	2:30-3:50pm
TITLE:	Curating a COVID-19 data repository and forecasting county-level death counts in the United States
SPEAKER:	**Bin Yu** Professor of Statistics, Electrical Engineering and Computer Sciences University of California, Berkeley

Abstract:
As the COVID-19 outbreak evolves, accurate forecasting continues to play an extremely important role in informing policy decisions. In this paper, we present our continuous curation of a large data repository containing COVID-19 information from a range of sources. We use this data to develop predictions and corresponding prediction intervals for the short-term trajectory of COVID-19 cumulative death counts at the county-level in the United States up to two weeks ahead. Using data from January 22 to June 20, 2020, we develop and combine multiple forecasts using ensembling techniques, resulting in an ensemble we refer to as Combined Linear and Exponential Predictors (CLEP). Our individual predictors include county-specific exponential and linear predictors, a shared exponential predictor that pools data together across counties, an expanded shared exponential predictor that uses data from neighboring counties, and a demographics-based shared exponential predictor. We use prediction errors from the past five days to assess the uncertainty of our death predictions, resulting in generally-applicable prediction intervals, Maximum (absolute) Error Prediction Intervals (MEPI). MEPI achieves a coverage rate of more than 94% when averaged across counties for predicting cumulative recorded death counts two weeks in the future. Our forecasts are currently being used by the non-profit organization, Response4Life, to determine the medical supply need for individual hospitals and have directly contributed to the distribution of medical supplies across the country. We hope that our forecasts and data repository at https://covidseverity.com can help guide necessary county-specific decision-making and help counties prepare for their continued fight against COVID-19.

Suggested Readings:
- "Curating a COVID-19 data repository and forecasting county-level death counts in the United States"
- "Perceptual Audio Coding Using Adaptive Pre and Post-Filters and Lossless Compression"
- "A tutorial on conformal prediction"