

Harvard Data Science Review • Special Issue 1 - COVID-19

Curating a COVID-19 Data Repository and Forecasting County-Level Death Counts in the United States

Nick Altieri, Rebecca L Barter, James Duncan, Raaz Dwivedi, Karl Kumbier, Xiao Li, Robert Netzorg, Briton Park, Chandan Singh, Yan Shuo Tan, Tiffany Tang, Yu Wang, Chao Zhang, Bin Yu

Published on: Nov 03, 2020

DOI: 10.1162/99608f92.1d4e0dae

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

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 23 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 nonprofit 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.

Just Accepted - Preview

11/3/20: To preview this content, click below for the **Just Accepted version** of the article. This peer-reviewed version has been accepted for its content and is currently being copyedited to conform with HDSR's style and formatting requirements.



[Altieri-et-al_Just-Accepted.pdf](#)

2 MB