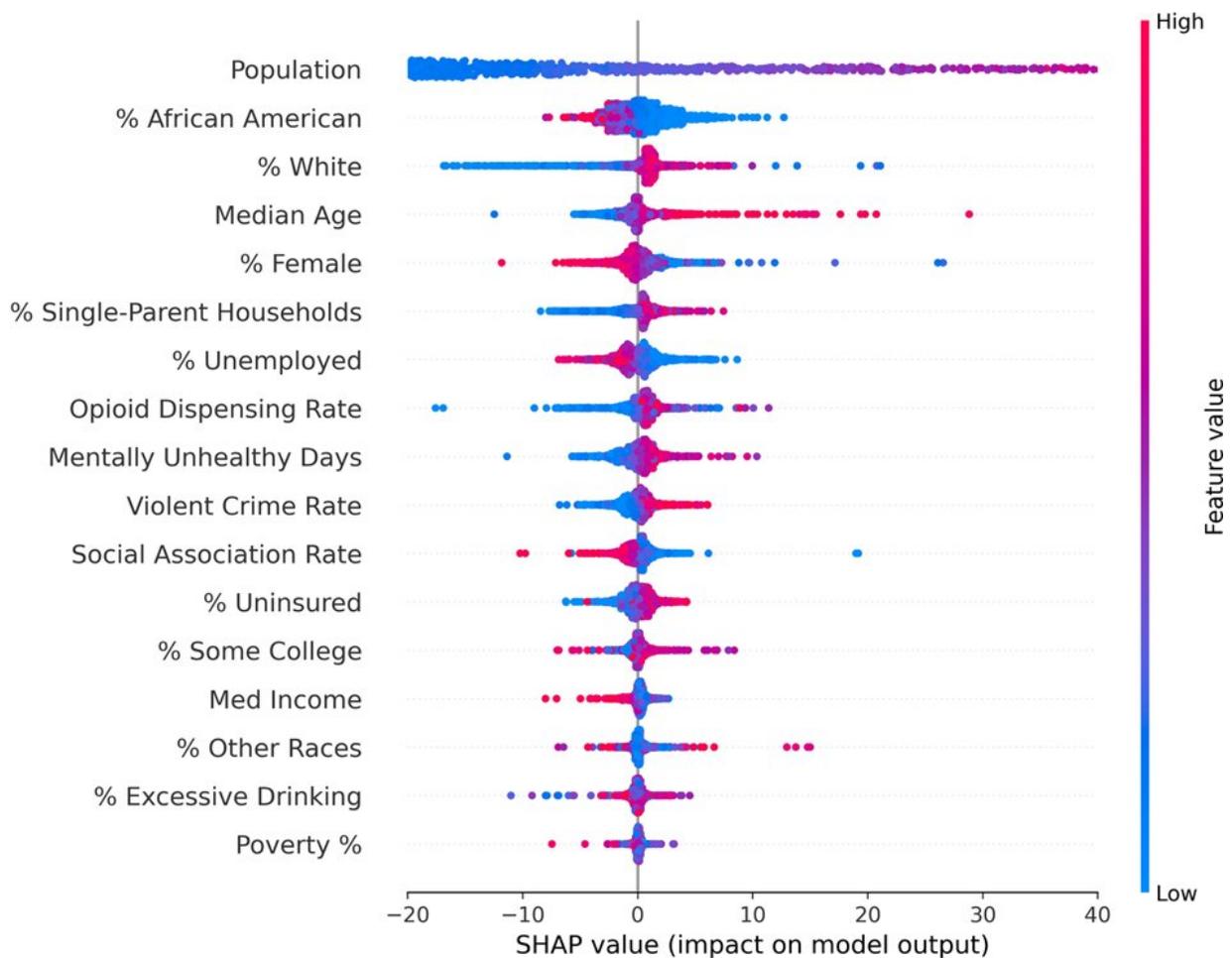


Suicide vulnerability index, machine learning model help predict counties' risk

June 28 2022, by Sarah Small



SHAP summary plot for the XGBoost-based suicide prediction model. SHAP summary plot reveals the relative impact of each feature on the prediction results. Credit: *npj Mental Health Research* (2022). DOI: 10.1038/s44184-022-00002-x

Suicide is a leading cause of death in the United States, but the models that have been used to predict suicide rates weight risk factors equally and rely on data for large geographic areas, limiting the precision of the predictions, according to Penn State researchers. Now, the researchers have developed a machine learning-based model that uses their newly developed suicide-vulnerability index, which weights risk factors, to identify at-risk communities at the U.S. county level.

The approach recently was published in *npj Mental Health Research*.

"Our goal was to develop a novel suicide-vulnerability index for U.S. counties with the help of a machine learning-based suicide prediction model," said paper co-author Soundar Kumara, Allen E. Pearce and Allen M. Pearce Professor of Industrial Engineering at Penn State, who is also affiliated with the College of Information Sciences and Technology. "By identifying the counties at higher risk for increased suicide rates, the model could help prompt targeted intervention programs."

The researchers analyzed data from 2010–19 at the county level in the 3,140 U.S. counties, the smallest possible geographic classification available in the Centers for Disease Control and Prevention's database. They identified 17 characteristics used to predict suicide rates and that could be categorized under demographics, socio-[economic factors](#) and health. The researchers suspected that some of these 17 characteristics would impact suicide rates more than others, and they set out to determine which factors affected suicide rates and by how much.

To identify the impact of each factor, the researchers used SHapley Additive exPlanations (SHAP), a game theory-based approach that explains how each variable contributes to the model's prediction.

"SHAP values explore the impact of each feature by comparing the

prediction results with and without that feature," said co-author of the paper Kristin Sznajder, assistant professor of public health sciences at Penn State College of Medicine, who is also affiliated with the Huck Institutes of the Life Sciences and the Population Research Institute. "Using the SHAP values, the importance of all 17 features used in the prediction model training set were identified. By identifying and isolating the top five important features from our analysis, we developed the suicide vulnerability index. In earlier work, such indexes were created by including all the variables without considering their effects on the output."

The top five county level features driving suicide-prediction results were population, percent African American population, percent [white population](#), median age, and percent female population, where higher population, percent white population and median age correlated with an increase in suicide rates while higher percent African American population and percent female [population](#) saw a decrease in [suicide rates](#)

Vishnu Kumar, Penn State industrial and manufacturing engineering graduate student and first author on the paper, emphasized that the SHAP values differentiated this [machine learning](#) prediction model from previous models.

"Several disciplines are using machine learning extensively to address data intensive problems," he said. "Machine learning models are often referred to as 'black boxes' because we don't know what is happening inside the models or the logic behind how the model compute results, even though their results are highly accurate. In this context, SHAP values provide a very convenient way to explain machine learning models and help us make powerful, fair and accurate interpretations and decisions."

The researchers said they hope that their work will lay the foundation for targeting and implementing suicide-intervention programs.

"An exciting opportunity for future work is to investigate the possibility of using machine learning techniques to gain insights to how variations in public health policies may affect rates of [suicide](#)," Sznajder said.

"Perhaps our model could be implemented at local and state levels to create early warning systems that can impact policy and resource allocation."

More information: Vishnu Kumar et al, Machine learning based suicide prediction and development of suicide vulnerability index for US counties, *npj Mental Health Research* (2022). [DOI: 10.1038/s44184-022-00002-x](#)

Provided by Pennsylvania State University

Citation: Suicide vulnerability index, machine learning model help predict counties' risk (2022, June 28) retrieved 4 May 2024 from <https://medicalxpress.com/news/2022-06-suicide-vulnerability-index-machine-counties.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.